

DOT/FAA/AR-01/96

Office of Aviation Research
Washington, D.C. 20591

A Methodology for the Assessment of the Capability of Inspection Systems for Detection of Subsurface Flaws in Aircraft Turbine Engine Components

September 2002

Final Report

This document is available to the U.S. public
through the National Technical Information
Service (NTIS), Springfield, Virginia 22161.



U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report. This document does not constitute FAA certification policy. Consult your local FAA aircraft certification office as to its use.

This report is available at the Federal Aviation Administration William J. Hughes Technical Center's Full-Text Technical Reports page: actlibrary.tc.faa.gov in Adobe Acrobat portable document format (PDF).

1. Report No. DOT/FAA/AR-01/96		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle A METHODOLOGY FOR THE ASSESSMENT OF THE CAPABILITY OF INSPECTION SYSTEMS FOR DETECTION OF SUBSURFACE FLAWS IN AIRCRAFT TURBINE ENGINE COMPONENTS				5. Report Date September 2002	
				6. Performing Organization Code	
7. Author(s) R.H. Burkel², C.-P. Chiou³, T.K. Keyes², W.Q. Meeker³, J.H. Rose³, D.J. Sturges², R.B. Thompson³, and W.T. Tucker²				8. Performing Organization Report No.	
9. Performing Organization Name and Address ¹Honeywell Engines, Systems & Services (AlliedSignal) 111 South 34th Street Phoenix, AZ 85072-2181 ³Iowa State University 1915 Scholl Road Ames, Iowa 50011		2General Electric One Neumann Way Cincinnati, OH 45212 General Electric Corporate Research & Development Center One Research Circle Schenectady, NY 12301 4Pratt & Whitney 400 Main Street East Hartford, CT 06108		10. Work Unit No. (TRAIS)	
				11. Contract or Grant No.	
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Aviation Administration Office of Aviation Research Washington, DC 20591				13. Type of Report and Period Covered Final Report	
				14. Sponsoring Agency Code ANE-110	
15. Supplementary Notes This is a report describing a portion of the efforts of Task C (Probability of Detection) of the Engine Titanium Consortium, Phase I with important input from C. Annis⁴, L.J. Brasche³, R.S. Gilmore², F.J. Margetan³, L. Schaeffer¹, K. Smith⁴. The FAA William J. Hughes Technical Center Technical Monitor was Mr. Bruce Fenton.					
16. Abstract A new methodology for determining probability of detection is described. Physical models are used heavily during the inspection process to minimize the amount of empirical data that must be gathered. This report includes a general review of various methodologies for determining probability of detection as well as a detailed discussion of the new approach that is being applied to the ultrasonic detection of internal inclusions in the rotating components of aircraft engines. Results of its application to the ultrasonic detection of flat-bottom holes and synthetic hard-alpha inclusions in laboratory measurements on flat plate samples are presented as well as a comparison to results obtained by other methodologies. This report summarizes the ongoing experiments and analysis aimed at validating the methodology of full geometry components in an industrial setting and extending its predictions to the detection of naturally occurring flaws. New tools will further reduce the need for empirical experiments through the use of physical models of microstructural effects on the ultrasonic response are included.					
17. Key Words POD, Nondestructive inspection, Turbine engine inspection			18. Distribution Statement This document is available to the public through the National Technical Information Service (NTIS) Springfield, Virginia 22161.		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 138	
22. Price					

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	xi
1. INTRODUCTION	1
2. OUTLINE OF THE REPORT	1
3. OBJECTIVES	2
4. MOTIVATION	3
4.1 Crash of United Airlines Flight 232 in Sioux City	3
4.2 Unique Issues Associated With Hard-Alpha Detection	4
4.3 Need for a New Methodology	6
5. APPROACH FOR A NEW METHODOLOGY	7
5.1 General S/N-Based Methodology Concept	7
5.1.1 Background: Hit-Miss Methods	7
5.1.2 Methods Using Signal Distributions	10
5.1.3 Methods Using Signal and Noise Distributions	12
5.1.4 Limitations of Existing Methods	14
5.2 ETC Implementation	15
5.2.1 Background	15
5.2.2 Distributions-of-Noise and Flaw-Plus-Noise Signals	19
5.2.3 General Approach	20
5.2.4 Fixed and Random Factors Affecting Flaw and Noise Distributions	21
5.2.5 Measurements of Responses of Nominally Identical Flaws	22
5.2.6 Modeling of the Generalized Signal/Prediction Deviations	25
5.2.7 Basic Probability of a True Detection (POTD)	28
5.2.8 POTD for Production Inspection	29
5.2.9 Probability of a False Alarm (PFA)	31
5.2.10 Relative Operating Characteristics (ROC)	31
5.2.11 Probability of Indication (POI)	32
5.2.12 Summary of Terms	32
6. SIGNAL MODELING AND VERIFICATION	33
7. RESULTS OBTAINED WITH THE NEW METHODOLOGY	34
7.1 Flat-Bottom Holes	35

7.2	Synthetic Hard-Alpha Inclusions	36
7.3	Comparison of Two Transducers	42
7.4	Uncertainty Bounds	42
8.	COMPARISON OF NEW METHODOLOGY PREDICTIONS FOR FBH POD TO PREDICTIONS OF OTHER METHODS OF ANALYSIS	43
8.1	Qualitative Response (Hit-Miss) Methods	44
8.1.1	American Society of Nondestructive Testing Recommended Practice	44
8.1.2	USAF/UDRI PF Program	46
8.1.3	Empirical ROC Approach	47
8.2	Quantitative Response Methods	47
8.2.1	USAF/UDRI \hat{a} Versus a Program	47
8.2.2	GEAE Effective Reflectivity Method	52
8.3	Conclusions	60
8.3.1	Comparison With New Methodology Results	60
8.3.2	Typical Sets of Ultrasonic Detectability Data	61
9.	PHYSICS-BASED DESCRIPTIONS OF NOISE AND SIGNAL-PLUS-NOISE DISTRIBUTIONS: THEORY, EXPERIMENTAL VERIFICATION, AND STRATEGIES FOR INCORPORATION IN THE NEW METHODOLOGY	61
9.1	Motivation	61
9.2	Empirical Studies of Noise Distributions	62
9.2.1	Mathematical Techniques	62
9.2.2	Experimental Techniques	64
9.2.3	Results	65
9.2.4	Conclusions	68
9.3	Model-Based Studies of Noise Distributions	68
9.3.1	Theoretical Models for Gated Peak-to-Peak Noise	68
9.3.2	Empirical Investigation of Gated Peak-to-Peak Noise	71
9.4	Modeling Signal-Plus-Noise Distributions	82
9.5	Beam Modulation Effects	88
9.6	Effects of Surface Roughness	88

9.6.1	Measurements of Surface Roughness	89
9.6.2	Theory for the Effects of Surface Roughness on Grain Noise	89
9.6.3	Theory for the Effects of Surface Roughness on the Signal From Subsurface Inclusions	89
9.6.4	Predictions of the Effects of Surface Roughness on the SNR	91
9.6.5	Summary of Crucial Results	93
9.6.6	Conclusions	94
9.7	Strategies for Incorporating Physics-Based Descriptions of Noise and Signal-Plus-Noise in the New Methodology	94
10.	FUTURE DIRECTIONS	95
10.1	Random Defect Block	95
10.2	Contaminated Billet Study	96
10.3	Portable POD	96
10.4	Adjusting for Natural-Flaw Parameters	96
11.	SUMMARY AND CONCLUSIONS	96
11.1	Summary	96
11.2	Conclusions	97
12.	REFERENCES	98
APPENDIX A—DETAILS OF FLAW SIGNAL MODELING AND VALIDATION		

LIST OF FIGURES

Figure	Page
1 Photomicrograph of Hard-Alpha Inclusion Showing Regions Embrittled by Enhanced Nitrogen Content (Light Microstructure) and Pores (Dark Regions) Produced by Fracture During Thermomechanical Processing	4
2 Metallographs of Typical Titanium Alloy Structure (a) Macrostructure (as Seen Under Crossed-Polarizer) and (b) Microstructure	5
3 Schematic Diagram Showing the Competition Between Backscattered Noise and Signals Reflected From Hard-Alpha Inclusions	5
4 Example of Rayleigh Distributions With Increasing Scale, Sigma	13
5 Example of Gaussian Distributions With Increasing Location (Mean) and Constant Scale (Standard Deviation)	14
6 Influence of Additional Information on Accuracy of a Linear Regression Fit to Data	16
7 A Schematic View of the Relationship Between POI and POTD	17
8 Examples of Possible Distribution of Noise Signals and Distributions of Flaw-Plus-Noise Signals, Showing POD and PFA by the Cross-Hatched Areas	19
9 Proposed ETC POD Estimation Methodology	21
10 C-Scans for Three of the Cases Considered in the FBH Factorial Experiment	23
11 C-Scan of SHA Block With 10-MHz Probe Focused in the Plane of the Flaws and Illuminating the Block at Normal Incidence	24
12 Plot of Data From the 10-MHz-Focused Transducer, Illuminating the 5.9% N SHA Block at Normal Incidence	25
13 Distribution of Beam Off-Center Deviations After $\lambda=0.3$ Transformation From the 10-MHz-Focused Transducer, Normal Incidence Scans of Cylindrical, Synthetic Hard-Alpha Inclusions	27
14 Schematic UT Scan Plan With Coarse Increments	30
15 Relative Operating Characteristic Curves	32
16 Basic POD (a) and ROC Functions for Two Different Sets of Inspection Conditions Using Transducer No. 2	36

17	Experimental Configuration for Studying Effect of Probe Tilt on POTD of SHAs	37
18	C-Scans of SHA Block With 10-MHz Probe Focused in the Plane of the Flaws and Illuminating the Block at a 5° Tilt and Four Azimuthal Orientations	38
19	Basic POTD for a 10-MHz Transducer at Normal Incidence, Positioned Directly Over, and Focused at the Same Depth as the Synthetic Hard-Alpha Flaw	39
20	POTD for a 10-MHz Transducer at Normal Incidence, Focused at the Same Depth as the Synthetic Hard-Alpha Flaw, Assuming 30-mil Scan Increments and a Very Narrow Gate Width	39
21	POTD for a 10-MHz Transducer at Normal Incidence, Focused at the Same Depth as the Synthetic Hard-Alpha Flaw, Assuming 0.060" (60-mil) Scan Increments and 0 μ s Gate Width	40
22	ROC for a 10-MHz Transducer at Normal Incidence, Focused at the Same Depth as the Synthetic Hard-Alpha Flaw, Assuming 0.030" (30-mil) Scan Increments	40
23	ROC for a 10-MHz Transducer at Normal Incidence, Focused at the Same Depth as the Synthetic Hard-Alpha Flaw, Assuming 0.060" (60-mil) Scan Increments	41
24	POTD for a 10-MHz Transducer at Normal Incidence, Assuming 0.030" (30-mil) Scan Increments and a Gate Width of 0.5 inch	41
25	Comparison of POI for Transducer No. 2 (5 MHz) and Transducer No. 4 (10 MHz) Using Normal Incidence, Assuming 0.020" (20-mil) Scan Increments and a Gate Width of 0.25 inch With the Threshold Adjusted Such That PFA = 0.02	42
26	POTD for a 10-MHz Transducer at Normal Incidence, Assuming 0.030" (30-mil) Scan Increments and a Gate Width of 0.5 inch, Showing Uncertainty Bands for ± 3 dB Uncertainty in the Model Predictions	43
27	American Society for Nondestructive Testing Recommended Practice POD for No. 1 FBHs at Two Confidence Levels as a Function of Threshold Level	45
28	USAF/UDRI PF Analysis POD for Planar Voids Based on PF Analysis (Transducer No. 2)	46
29	Comparison of Scatter-Plots of Signal Response for Transducer No. 2 Versus FBH Area or FBH Diameter in Linear or Logarithmic Coordinates	49
30	Comparison of Transducer No. 2 Data With \hat{a} Regression Line	49
31	Cumulative Normal Probability Plot of the Residuals From the \hat{a} Regression of Log (mV) on Log (Area) for Transducer No. 2	50

32	USAF/UDRI \hat{a} Analysis of POD vs Flaw Area for Planar Voids	51
33	USAF/UDRI \hat{a} Analysis of POD vs Flaw Diameter for Planar Voids	51
34	Comparison of Data With R_e Regression Line (Transducer No. 2)	53
35	Cumulative Normal Probability Plots for R_e and Log (R_e) (for Transducer No. 2)	54
36	GE Effective Reflectivity Analysis POD for Planar Voids, Linear Area-Amplitude Model (Transducer No. 2)	55
37	Comparison of Noise-Adjusted Data With R_e Regression Line (for Transducer No. 2)	56
38	Cumulative Normal Probability Plots for Noise-Adjusted R_e and Log (R_e) (for Transducer No. 2)	56
39	GE Effective Reflectivity Analysis POD for Planar Voids, Noise-Adjusted Linear Area-Amplitude Model (Transducer No. 2)	57
40	Comparison of ISU Model With \hat{a} and R_e Regression Lines	58
41	Cumulative Normal Probability Plots for the ISU Model (Transducer No. 2)	58
42	GE Effective Reflectivity Analysis POD for Planar Voids, ISU Nonlinear Area-Amplitude Model (Transducer No. 2)	59
43	Revised GE Effective Reflectivity Analysis POD for Planar Voids, ISU Nonlinear Area-Amplitude Model (Transducer No. 2)	60
44	Plots of Estimated and Experimental Distributions for a 5-MHz, 0.100" Diameter Transducer on a Ti-17 Test Block (a) Lev Fit With $\mu = 33.405$, $\sigma = 4.221$, rms = 0.003; (b) Log Norm Fit With $\mu = 3.567$, $\sigma = 0.135$, rms = 0.002; and (c) Weibull Fit With $\mu = 35.728$, $\sigma = 4.931$, rms = 0.009	66
45	The Low Probability or Tail Area of the Experimental and Estimated PDFs Derived From Grain Noise Data Collected With a 5-MHz, 0.100" Beam Diameter Transducer on a Ti-17 Sample Block	67
46	Random Flights Model	70
47	Fitting Various Distributions to Noise Data From the High Noise Region of the Contaminated Billet	75
48	Fitting Various Distributions to Noise Data From a Forging	77

49	Schematic of Noise Variations in a Line Scan Through Flaw (a) Correlation Length of Noise on the Order of Flaw Response and (b) Correlation Length of Noise Small With Respect to Flaw Response	83
50	Experimental Results of Line Scans Through No. 2 SHAs	83
51	Rician Distribution for Various Values of The Signal-to-Noise Ratio $A^2/2\sigma^2$	85
52	Synthetic Flaw Signal (a) Noise Free and (b) $A^2/2\sigma^2 = 4$	86
53	A Comparison of Simulated Distributions of Gated Peak-to-Peak, Signal-Plus-Noise Distributions to Physically Based Model Predictions in Ti 6-4, Sample K1	87
54	A Comparison of the Discrete Responses of Eight SHAs to the Physically Based Model Predictions of the Distribution	88
55	Normalized SNR (NSNR) as a Function of Depth for Different Radii Inclusions ($A = 0.5, 1.0, \text{ and } 2.0 \text{ mm}$) for a Focused Transducer	92
56	Normalized SNR as a Function of Depth for Different Transducer Radii ($R = 10, 20, \text{ and } 40 \text{ mm}$) for a Probe Focused at 40 mm Inside the Metal	92
57	Strategy of Incorporating Physics-Based Descriptions of Signal and Noise Distributions in POD Methodology	95

LIST OF TABLES

Table		Page
1	Issues in the Detection of Hard-Alpha Inclusions and Subsequent Life Prediction That Lead to the Need for a New POD Methodology	7
2	Probability of Detecting No. 1 FBHs With a 400-mV Threshold	45
3	Probability of Detecting No. 3 FBHs With Transducer No. 2	47
4	Minitab Release 10.5 <i>Xtra</i> Tests Applied to \hat{a} Regression for Transducer No. 2	48
5	Hole Size Exhibiting 50% POD for Several Methodologies as a Function of Threshold	61
6	Likelihood Modifications for Censoring	73
7	Statistical Measures for Billet Noise Data	79

8	Statistical Measures for Forging Noise Data Obtained With a Transducer Have a Nominal Center Frequency of 5 MHz and Using a 4- μ sec Time Gate	79
9	Comparison of Values of N Computed From Nominal Probe Parameters to Those Determined by Maximum Likelihood Analysis of Forging Noise Data	81
10	Optimum Frequency for Detection of Worst-Case Hard-Alpha Inclusions Beneath Smooth and Rough Surfaces	90

LIST OF ACRONYMS AND SYMBOLS

BCF	Beam-centered flaws
CBS	Contaminated Billet Study
CDF	Cumulative distribution function
ESWD	Equivalent square wave duration
ETC	Engine Titanium Consortium
FBH	Flat-bottom holes
FKIR	Finite-beam Kirchhoff
FOM	Figure of merit
GEAE	General electric aircraft engines
GECRD	General Electric Corporate Research and Development
GOF	Goodness-of-fit
Ka	1
IID	Independent and identically distributed
LEV	Largest Extreme Value
MOOT	Method of Optimal Truncation
MLE	Maximum likelihood estimation
PDF	Probability density function
PFA	Probability of false alarms
PKIR	Plane-wave Kirchhoff model
POI	Probability of an Indication
POD	Probability of detection
POTD	Probability of True Detection
RF	Radio-frequency
ROC	Relative Operating Characteristics
RMS	Root mean squared
RV	Random variable
SHA	Synthetic hard-alpha inclusions
SNR	Signal-to-noise ratio
UT	Ultrasonics

EXECUTIVE SUMMARY

This report describes a portion of Task C—Probability of Detection of Phase I of the Engine Titanium Consortium. In particular, the development of a new methodology for the determination of probability of detection (POD) is reported. It is then applied to the estimation of the POD for the ultrasonic detection of flat-bottom holes (FBHs) and synthetic hard-alpha inclusions (SHAs) in aircraft engine titanium alloys.

The report opens with a discussion of the background that motivated this research. This research encompasses the following:

1. The Sioux City incident.
2. The need to determine the POD for the ultrasonic detection of the internal, hard-alpha inclusions that were its cause.
3. The existing methodologies do not provide a comprehensive determination of POD for the ultrasonic detection of such naturally occurring internal defects.

In a detailed discussion of this background, the statistical detection theory, based on a determination of signal and noise distributions that exist with and without the presence of a flaw, shows much promise in providing the required answers. It is further noted that this approach would have the additional advantage of quantifying the probability of false alarms (PFA). This would provide a rational basis for considering the quantitative tradeoffs between high-detectability (safety related) and high-reject levels (cost related). Because of the rarity of the natural hard-alpha inclusions, the use of physical models of the flaw response, which limits the amount of experimental information required, holds particular promise. A methodology that relies on such models would have the derived benefit of allowing estimates of POD to be made in new situations. However, samples are not available because of cost, time, or fabrication limitations. This report describes the methodology that was developed in response to these motivations.

This report contains five major sections. Sections 1-4 develop the objectives and motivation in more detail. Section 5 presents the new methodology as well as a detailed discussion of the underlying physical and statistical assumptions. Motivated by the considerations summarized above, this methodology uses the physical models of the inspection process to predict what the flaw response would be in the absence of multiple sources of variation such as microstructural effects. Section 6, supported by appendix A, discusses these models and a preliminary set of validation experiments. These experiments are based on simple geometry and test conditions that have been conducted to determine their accuracy for these limited test cases. Section 7 presents examples of POD predictions for FBHs and SHAs that have been made using the new methodology, based on conditions that exist in the laboratory examination of flat plates. Included are predictions of the effects of such experimentally controllable factors as scan plan, gate width, and transducer frequency on POD. Further verification is described in section 7, in which these predictions of the new methodology are compared to those of existent methodologies for the case of FBHs. The need for the new methodology was reconfirmed

because the existing methodologies were unable to make similar comparative predictions for the SHAs.

Section 8 provides a comparison of new methodology predictions for FBH POD to predictions of other methods of analysis. Section 9 presents the results for studies of the functional forms of signal and noise distributions. These results will be included in future implementations of the new methodology. The goal of this work will be to incorporate as much physical understanding about the forms of these distributions into the methodology. This will further reduce the amount of experimental data that is required and increase the accuracy in treating issues such as the large tail of the noise distribution that controls false rejects.

Section 10 indicates the nature of in-progress work, including further verification for full-scale components under industrial inspection condition and extension to naturally occurring hard-alpha inclusions.

Those studies, which were completed after the work described herein but before the publication of this report, demonstrated that this methodology will require further refinement in order to accurately predict the detection probabilities of defects in a realistic industrial environment.

In particular, they show the need to “productionize” the methodology taking into account a number of input parameters to the physics-based models are not fully controlled in an industrial environment and thus represent a source of additional variability. Follow-on work is planned to develop approaches to introduce these and any other sources of variability into a POD prediction that is full representative of the industrial setting.

Future reports will document the results of those studies upon their completion. Section 11 summarizes the current status of the POD effort and provides some concluding comments.

1. INTRODUCTION.

The probability of detection (POD) function relates the detectability of a flaw to its size and other features. It quantifies the ability of a nondestructive measurement system to detect flaws and provides a means to make improvements in inspections given changes in measurement techniques, procedures, and systems. This metric has a variety of uses. From one perspective, it provides a basis for assuring the public and/or customers that adequate detectability has been achieved in particular applications. From another perspective, the POD provides the link between the inspection and design communities. Thus, this link is an integral part of life management programs. In the context of the latter applications, the POD must be combined with other information (such as “*a priori*” information about defect distributions) to make a prediction of the distribution of expected part lifetimes, as influenced by inspection.

This report describes a new methodology for determining the POD and its application to subsurface flaws in aircraft engine components. A renewed attention to the detection of subsurface flaws was motivated by a serious accident in 1989 that was attributed to hard-alpha inclusions in titanium alloys. Given known limitations of existing methodologies in evaluating the POD of ultrasonic techniques that were designed to detect these hard-alpha inclusions, the development of a new methodology was undertaken as one task of the Engine Titanium Consortium (ETC). As the methodology developed, it provided a number of other advantages beyond those initially motivating the work. This report documents the development of the methodology, its early applications and verification against existing methodologies, as well as future steps that are planned for further development and utilization.

2. OUTLINE OF THE REPORT.

The first four sections of this report provide background and motivation. The major technical results will be presented in sections 5-9. Section 5 presents the technical approach developed by the ETC. In section 5.1, the conceptual basis of the approach, the application of statistical detection theory augmented by physical models of the inspection process, is presented. In essence, one seeks to explicitly develop a quantitative description of the distributions of signal and noise that govern the detection process using well-validated physical models wherever possible to limit the number of samples that must be prepared and experiments that must be conducted. In section 5.2, the general form of the implementation of these ideas, which has been developed by the ETC, is presented.

Section 6 summarizes the physical models that are incorporated in the methodology, including the physical approximations made and the results of experimental verifications. Detailed information to support that summary is deferred to appendix A to assure that extended discussions of ultrasonic theory and experiment do not detract from the continuity of this report, which is otherwise concerned with issues of statistical reliability analysis.

Section 7 presents the results of applying this methodology to two classes of defects, flat-bottom holes (FBHs) and synthetic hard-alpha (SHAs) inclusions in aircraft engine titanium alloys. Included are predictions of POD, probability of false alarms (PFA), and their combination in the

form of relative operating characteristics (ROC) as influenced by details of measurements systems, e.g., frequency, transducer characteristics, and scan plans.

In section 8, existent methodologies are applied to the same set of data to predict the POD of FBHs. The comparisons of the results of sections 7 and 8 serve both to illustrate some of the difficulties in applying existent methodologies to internal flaw data, and when these have been overcome, to verify the predictions of the new methodology.

In the course of the activities of the POD task and its coupling to other parts of the ETC such as the Fundamental Studies Task, information was developed that time did not allow to be incorporated in the first-generation methodology but which will be available for the next generation. Section 9 documents the understanding which deals with the proper functional forms of the statistical descriptions of signal and noise distributions, the effects of surface roughness, and material effects which modulate the beam amplitude and profile.

Section 10 provides a discussion of future directions. Included are brief descriptions of the Random Defect Block experiment, which provides an opportunity to validate the physical models and the methodology against representative production configuration, and of the Contaminated Billet Study, in which ultrasonic scattering from naturally occurring, hard-alpha inclusions is being carefully measured, after which the defects are being successively sectioned to obtain complete flaw descriptions. These results will be used to incorporate information from naturally occurring hard-alpha inclusions in the POD predictions of the methodology. A discussion follows of the methodology to obtain a “portable POD” which realizes some of the desirable attributes discussed at the end of section 4, e.g., the estimates of the effects of changes in inspection procedures for engineering and management purposes. The section concludes with a discussion of strategies for adjusting the methodology for the parameters of naturally occurring flaws.

Section 11, summarizing what has been accomplished and issues that remain outstanding, completes the report.

3. OBJECTIVES.

The objectives of the POD Task of the ETC was to develop a methodology to:

- determine the POD of flaws, especially subsurface, in materials.
- verify the methodology.
- provide appropriate quantitative information to allow risk and life management studies to be carried out.

This report describes the development of the methodology and its ability to verify flat-bottom holes and synthetic hard-alpha inclusions in titanium alloys.

4. MOTIVATION.

4.1 CRASH OF UNITED AIRLINES FLIGHT 232 IN SIOUX CITY.

Quantification of the nondestructive POD inspections is an integral part of ensuring initial safety and managing safe life of critical structural components. For the case of titanium rotating components of aircraft engines, attention was focused on such inspections and their reliability by the 1989 crash of United Airlines Flight 232 in Sioux City, Iowa. Despite the heroic efforts of the pilot and crew, 111 lives were lost in the crash. The cause of the crash was found to be the separation of the stage 1 fan disk resulting in the loss of hydraulic fluid and control of the airplane. Subsequent analysis revealed that this disk separation was the result of a fatigue crack that had originated from a metallurgical inhomogeneity (i.e., a hard-alpha inclusion) which consisted of a region embrittled by enhanced nitrogen content and contained microporosity and microcracks. The hard-alpha inclusion had gone undetected during the disk manufacturing process inspections. Also, neither the inclusion nor the resulting fatigue had been detected during in-service inspections.

This incident led to an enhanced concern regarding the reliability of the inspection of aircraft engine components, particularly those made from titanium alloys. In the Titanium Rotating Components Review Team Report [1], a number of recommendations were made regarding improvements in inspection and its quantification. Included in the latter were the following.

- Manufacturing Inspection. “Require the highest standard (smallest flat-bottomed hole (FBH) or equivalent) practicable in the industry for the size of part being inspected.”
- In-Service Inspection. “Develop criteria, within two years, to inspect all critical, life-limited, in-service parts at intervals established by fracture mechanics technology.”
- Design Procedure. “Require life management methodologies to consider the effect of metallurgical defects on part life, accounting for the maximum defect sizes which may be missed during production and in-service inspections.”
- Research and Development. “Establish industry wide probability of detection (POD) curves for fluorescent penetrant, ultrasonic, and eddy-current manufacturing and in-service inspection methods and processes.”
- FAA Policy and Guidance. “Develop new advisory material on lifing analysis and life management procedures for engine life-limited parts.”

Each of these explicitly or implicitly identifies the need for a quantification of the POD of the inspection techniques and illustrates the importance of that quantity in life management.

In response to the Sioux City accident and the recommendation of the Titanium Rotating Components Review Team Report, the FAA formed the Engine Titanium Consortium, consisting of the closely coupled efforts of AlliedSignal, General Electric, Pratt & Whitney, and Iowa State University in a consortium facilitated by Iowa State University. The Phase I program of the ETC

had four tasks: Fundamental Studies in Titanium, Ultrasonic Inspection in Production, Eddy-Current Inspection In-Service, and Probability of Detection. This report documents a portion of the efforts of the Probability of Detection Task.

4.2 UNIQUE ISSUES ASSOCIATED WITH HARD-ALPHA DETECTION.

Two factors make it difficult to detect hard-alpha inclusions. First, these distributed defects can have a complex structure and associated ultrasonic reflections can be quite weak. The defects are caused by the presence of excess nitrogen or oxygen that is occasionally introduced into the material during ingot preparation due to a variety of possible causes [1]. These impurities occupy interstitial sites in the lattice, leading to a brittle structure that is prone to cracking. Figure 1 shows a micrograph of a hard-alpha inclusion. The lighter region is the microstructure where enlarged, brittle, alpha-phase grains have been created by the presence of nitrogen.

It is difficult to detect this condition due to the presence of interstitial impurities that cause the density and elastic constants of the embrittled region to differ slightly from those of the titanium alloy in which they are embedded. With typical variations in density and elastic constants, the properties, which control ultrasonic reflectivity, might be on the order of 10% [2]. Moreover, since there is a gradient in nitrogen content, the reflections are somewhat diffuse. Hence, hard-alpha inclusions will produce very weak ultrasonic reflections, which can be difficult to detect. This difficulty is often partially mitigated by the presence of small pores or cracks, indicated by the dark regions in figure 1. This may be created during the thermomechanical processing of the material by the fracture of the brittle hard-alpha inclusions. Although these pores or cracks are stronger reflectors than the embrittled titanium, they may be small and have complex morphologies. They therefore can produce relatively weak signals themselves.

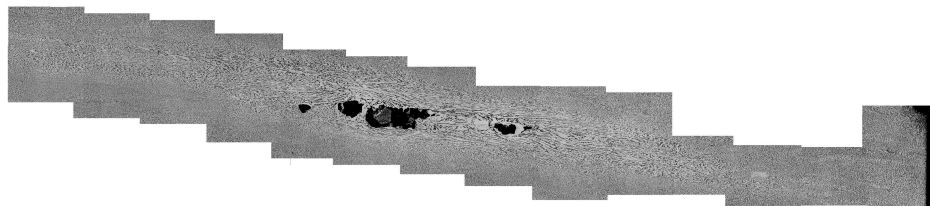


FIGURE 1. PHOTOMICROGRAPH OF HARD-ALPHA INCLUSION SHOWING REGIONS EMBRITTLLED BY ENHANCED NITROGEN CONTENT (LIGHT MICROSTRUCTURE) AND PORES (DARK REGIONS) PRODUCED BY FRACTURE DURING THERMOMECHANICAL PROCESSING

Secondly, the detection of these defects is even more difficult because of the structure of titanium alloys. Aircraft engine alloys have a two-phase structure that can be very complex, including features at both the microstructural (having dimensions on the order of micrometers) and macrostructural (having dimensions on the order of millimeters) scales. Figure 2 illustrates this by showing etched and illuminated metallographs that reveal these structures. As a result, the material can appear quite inhomogeneous to an ultrasonic wave. Hence, as the ultrasonic wave propagates through the material, there is much backscattered noise arising from benign

inhomogeneities in the microstructure. This tends to mask the weak signals from the hard-alpha inclusions. Figure 3 schematically summarizes the problem.

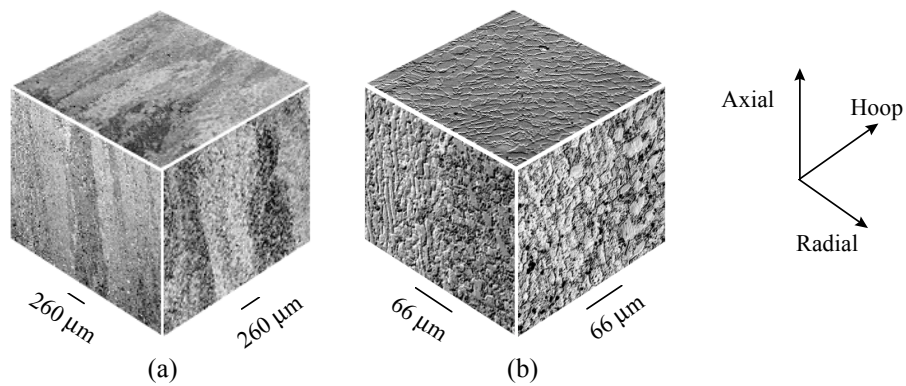


FIGURE 2. METALLOGRAPHS OF TYPICAL TITANIUM ALLOY STRUCTURE
(a) MACROSTRUCTURE (AS SEEN UNDER CROSSED-POLARIZER) AND
(b) MICROSTRUCTURE

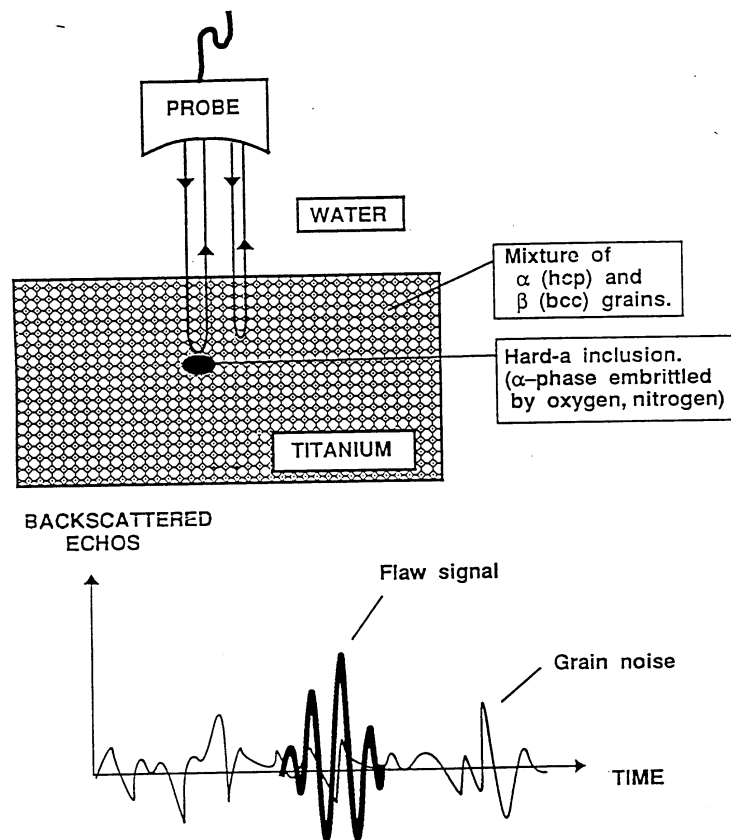


FIGURE 3. SCHEMATIC DIAGRAM SHOWING THE COMPETITION
BETWEEN BACKSCATTERED NOISE AND SIGNALS REFLECTED
FROM HARD-ALPHA INCLUSIONS

The situation is further complicated by other effects that the macrostructure has on the ultrasonic beam itself. Included are steering of the beam away from the intended direction and the development of phase fluctuations across the beam [3]. The latter can weaken the strength of signals returned from what would normally be a strong reflector such as a planar surface or flat-bottom hole.

4.3 NEED FOR A NEW METHODOLOGY.

A number of empirical methodologies already exist which quantifies POD. In general, these are based on the idea of presenting to the inspector or inspection system a series of opportunities that are used to define the performance of the inspection. A detailed discussion of some of these methods can be found in reference 4. In earlier techniques, a number of opportunities are presented in each of several size ranges, yet the only information that is recorded is whether or not the flaw had been detected. POD and associated confidence levels are determined by performing a statistical analysis of the results.

The signal response approach that was developed later records the amplitude in each inspection. Certain assumptions are made about the shape the distribution signals taken from nominally identical flaws, and the data are used to determine the parameters of that distribution. This procedure allows POD information to be obtained from a significantly smaller number of samples. Further discussion of these techniques are found in section 5.1. Section 8, on the other hand, provides a preliminary verification of the new methodology by comparing its predictions, using existing methodologies, to obtain the same data for the flat-bottom hole case in titanium alloys.

When existing methodologies were applied to the ultrasonic detection of complex, subsurface defects such as hard-alpha inclusions, severe limitations were experienced. Table 1 summarizes some of the issues in contrast to the situation for conventional inspection problems, such as the detection of surface-breaking, low-cycle fatigue cracks, and life prediction based on the results. In either class of problems, such conditions as ultrasonic material properties, component geometry, and inspection parameters are generally known or can be obtained by simple measurement. However, whereas flaw parameters can be obtained by direct observations for relatively simple cases, it is difficult through direct observation to detect low-cycle fatigue cracks for internal hard-alpha inclusions. Moreover, if one wishes to develop a correlation between flaw response and flaw parameters, the inability to replicate specimens with flaws that represent those that occur naturally is a further problem, as well as the absence of a nondestructive referee technique. Finally, when relating the flaw parameters to structural severity, both the difficulty in producing representative flaws and a multitude of parameters (variable nitrogen content and distribution and complex pattern of pores and cracks), which pose important challenges, must be considered.

TABLE 1. ISSUES IN THE DETECTION OF HARD-ALPHA INCLUSIONS
AND SUBSEQUENT LIFE PREDICTION THAT LEAD TO THE
NEED FOR A NEW POD METHODOLOGY

Condition of Interest	Availability for More Conventional Problems	Availability for Hard-Alpha Analysis
Ultrasonic material properties-base metal	Material property tests	Material property tests
Geometry of components	Direct observation	Direct observation
Inspection parameters	Direct observation	Direct observation
Flaw parameters	Direct observation	Direct observations difficult or impossible
Correlation of response to flaw parameters	From destructive or nondestructive evaluation	Limited by inability to produce specimens with representative flaws, no non-destructive referee technique available
Correlation of flaw parameters to structural severity	Material/structural testing	Too many parameters for complete severity assessment, flaws difficult to simulate

Another motivation for seeking a new methodology is the fact that existing methodologies do not quantify the PFA. This is extremely important to the original equipment manufacturers (OEMs) since economic losses associated with false calls play an important role in the establishment and utilization of inspection technologies. The choice of threshold, which determines POD for a given inspection, will be strongly influenced by the PFA that is acceptable and also determined by that threshold.

Given the need for a new methodology driven by these issues, other desirable attributes have been identified. Included are the ability to use information from all sources, e.g., laboratory studies and the analysis of signals from naturally occurring defects found in the field; the ability to provide estimates of the effects of changes in inspection procedure on POD without requiring a new set of samples and experiments each time; and the ability to rapidly use that information to provide engineering feedback to nondestructive evaluation (NDE) personnel, management, and the lifing community.

5. APPROACH FOR A NEW METHODOLOGY.

5.1 GENERAL S/N-BASED METHODOLOGY CONCEPT.

5.1.1 Background: Hit-Miss Methods.

In the great majority of NDE applications, the flaw detection capability of specific techniques has been expressed indirectly through a reference to the size of the artificial or simulated flaws. These techniques are used to ensure that the sensitivity of inspections is under control.

For example, small notches and fatigue cracks in surrogate geometries are used to establish eddy-current inspections that are conducted at the same sensitivity, independent of when, where, or by whom the inspection is carried out. Flat-bottom holes (which are effectively planar, circular, subsurface voids) serve the same purpose for ultrasonic inspections.

This type of sensitivity standardization (sometimes called “calibration”), while effective for process control purposes, does not address the capability of the NDE technique to detect “real” (i.e., naturally occurring) flaws. The size of the reference flaw—the notch or FBH—provides some indication about the approximate size of the crack or inclusion that might be detectable. However, this is only a very approximate estimate. The properties of naturally occurring flaws are unlikely to coincide with those of the reference flaws. Because the differences in these properties are unclear, it is difficult to make more than a qualitative statement about their relative detectability. For example, it is possible to forecast that inclusions will be less detectable ultrasonically than FBHs of the same size or that doubling the scan increment is likely to decrease the likelihood of detecting small flaws. Therefore, it is much more difficult to determine how large the difference in detectability might be.

One of the earliest attempts to study this relationship was launched in the power generation industry in the late 1950s. Flaws that had been detected by ultrasonic inspection of steam turbine and generator rotor forgings were trepanned out. Their sizes were compared with corresponding estimates based on the FBH’s sizes used in standardizing the various inspections (e.g., using a simple mathematical model for the dependence of FBH response on size and distance from the transducer and the assumption that the response from natural flaws would behave similarly). This concept of the “Equivalent Flat-Bottom Hole” (EFBH)—the size of the FBH, at the same depth as the natural flaw, would give the same predicted signal as the natural flaw—allowed an extensive database to be developed, which compared the measured flaw sizes with the EFBH sizes. These results were used to estimate the residual life of other rotors containing ultrasonic indications [5]. However, they were not expressed in probabilistic terms.

Probabilistic concepts of flaw detectability were developed by the aerospace industry in the late 1960s. The motivation behind the concepts were to provide data for calculations to determine the life of cyclically stressed components by using fracture mechanics techniques. One of the inputs to such programs was the estimate of flaw sizes that might be in a component as a result of the properties of the materials, processes used in its manufacture, and that might remain in it even after NDE. It is a given that NDE techniques cannot guarantee that flaws of all sizes will be able to be detected. One of the inputs to such programs was the estimate of flaw sizes that might be in a component as a result of the properties of the materials and processes used in its manufacture and that might remain in it even after NDE. However, it has been found that NDE techniques cannot guarantee the detection of all flaw sizes. Expressing detectability as a function of flaw size has proven useful. However, the number of factors that could affect detectability was much larger than the independent pieces of information that were provided by the NDE technique. Thus, detectability was expressed in probabilistic terms as a probability of detection.

As an example of this, consider ultrasonic inspection that typically identifies only the amplitude and time delay of response signals (indications). However, the response is affected by the shape, orientation, and chemical-physical character of the flaw as well as by its size and depth. These additional factors result in a scatter in the response from flaws of a specific size. This scatter may be thought of as random in nature, which leads in turn to the concept of POD.

The U.S. Air Force's interest in fracture mechanics techniques began in the late 1960s and was initially focused on applications to airframes. It led to the sponsorship of several major data acquisition programs and the publication of MIL-STD-1530A (USAF) in 1975, which described requirements for the Aircraft Structural Integrity Program (ASIP). Contemporaneously, the National Aeronautics and Space Administration began awarding contracts for acquisition of quantitative NDE data and published a Shuttle Fracture Control Plan in 1974. The results from these studies were reported in terms of POD graphs as a function of size for various combinations of process, material, inspector, etc. The 1970s saw the use of POD concepts by individual industrial companies for the comparison of the effectiveness of inspection procedures.

As mentioned in section 3.3, the methods for POD analysis that were in use typically treated all inspection data as binomial, regardless of whether the NDE technique produced a response that was proportional to the size of the flaw. However, all that was recorded was whether or not the response from an individual flaw exceeded some predetermined detection threshold, namely, a "hit" or a "miss." Data acquisition obviously required prior knowledge of the existence and approximate size of each flaw. This requirement focused attention on evaluation of techniques for detection of surface-connected flaws such as the low-cycle fatigue (LCF) cracks. Methods were developed for generating such flaws under controlled conditions, usually by using notches to start crack growth as the samples were cyclically stressed. Once crack growth was initiated, the notches were machined off, and the crack was allowed to grow until the desired crack length was attained as indicated by using a high-powered optical microscope.

One of these binomial methods of analysis was published as a Recommended Practice by the American Society for Nondestructive Testing [6]. It involved using a moderate number of flaws, preferably of the same nominal size. If inspection of seven flaws resulted in all seven being detected, a POD of 90% with 50% associated confidence, may be claimed for this size and number. Detection of 29 out of 29 (or 45 out of 46, or 59 out of 61, etc.) is needed to establish 90% POD with 95% confidence. By repeating this approach with flaws of various sizes, the approximate dependence of POD on flaw size may be established. This is rarely attempted, however, since the number of flaws required may be daunting.

Other methods of analysis used in the 1970s and early 1980s tried to demonstrate the functional dependence of POD on flaw size with fewer flaw samples. This was done by using a variety of arbitrary averaging techniques, which are less rigorous statistically. One example was the Range-Interval Method (RIM) in which flaws were grouped into adjacent flaw size ranges, and the simple proportion of flaws detected in each range were treated as the best estimate for the true POD for flaws in that range. Graphs were plotted of these POD estimates versus the medium flaw size (or, sometimes, the smaller flaw) of the corresponding range. The individual data points frequently failed to show the monotonically increasing dependence of POD on size that was expected on intuitive or physical grounds. This deficiency was usually remedied by

drawing a smooth monotonically increasing curve through the general vicinity of the plotted data. The obvious subjectivity involved in this procedure led to several widely used alternatives that tried to generate a smooth monotonic curve by resorting to a variety of data-averaging techniques. One example of this approach is the Floating 60-Point Averaging Method. In it, the POD was first calculated for the largest 60 flaws. For the second POD calculation, they were omitted and replaced by the 61st largest flaw and so on until a POD had been calculated for the smallest 60 flaws. This approach was objective because the result obtained was independent of the analyst and produced curves that approximated—but still did not quite achieve—the desired monotonically increasing form.

The Pass/Fail (PF) option in the POD software [7] developed under USAF sponsorship was a culmination of the development of methods for analyzing hit/miss data. The PF program initially assumed¹ a log-logistic model² for the dependence of POD on flaw size, viz:

$$POD(a) = \frac{\exp(\alpha + \beta \cdot a)}{1 + \exp(\alpha + \beta \cdot a)} \quad (1)$$

where “ a ” is the flaw size and α and β are parameters of the model, the values of which are determined by maximum-likelihood estimation for each set of inspection data. Provided that there is an overlap between the sizes of detected and nondetected flaws, the PF program usually³ generates graphs of both the POD estimate and the associated lower one-sided 95% confidence bound on POD, as functions of flaw size, that show the expected smooth monotonically increasing dependence. The PF method has been widely applied to the estimation of POD for penetrant processes, for which the inspection data have typically been recorded only in pass/fail form, even though some approximate measurement of indication size is an intrinsic part of many such inspection procedures.

5.1.2 Methods Using Signal Distributions.

The hit/miss decision in an NDE process is based on signal response data that depends on various parameters, including the flaw size. Upon completing the development of POD analytical methods such as the PF program, it was recognized that response data generates more information than simply the hit-miss data. Consequently, an alternative method of analysis was developed to deal with response data. Since response data may be treated as a measure (albeit frequently imprecise) of the flaw size producing it, statistical conventions resulted in the response being designated as \hat{a} , and the POD method [8]. Thus, it became known as \hat{a} versus a ⁴.

Since eddy-current instruments produce such data in readily interpretable electronic form that are compatible with the same low-cycle fatigue cracks that were being generated for use with the earlier hit/miss methods, the \hat{a} versus a method has been widely used in the analysis of eddy-current data. This application was encouraged by the extension of fracture mechanics techniques

¹ This “assumption” was based on empirical observation that it provided a better fit to experimental data than any of several competing models under review.

² In recent versions of the PF program, the log-logistic model has been replaced with a cumulative lognormal model.

³ The calculation sometimes fails to converge.

⁴ Read as \hat{a} versus a .

to aircraft engine applications with requirements defined by MIL-STD-1783 (USAF) that was published in 1984.

As in the PF method, a parametric approach is used where the dependence of POD on flaw size is modeled⁵ by a cumulative lognormal distribution function (Φ):

$$POD(a) = 1 - \frac{\Phi[\ln(a_{dec}) - (\beta_0 + \beta_1 \cdot \ln(a))]}{\sigma} \quad (2)$$

where a_{dec} is the decision threshold used in deciding whether an indication is classified as detected or not. The parameters β_0 and β_1 of this function are obtained from regression analysis of the dependence of $\ln(\hat{a})$ on $\ln(a)$. The third regression parameter, σ , is the standard deviation of the distribution of the responses for any given flaw size (which usually can be shown to be plausibly normal in these \ln - \ln coordinates). The \hat{a} versus a method is thus able to take advantage of the well-documented properties of the normal distribution in extracting estimates of $POD(a)$ from the $\ln(\hat{a})$ versus $\ln(a)$ data. As with the PF method, it generates curves showing the dependence on flaw size of the mean POD and of the associated lower one-sided 95% confidence limit. As with the PF method, it is typically applied to laboratory-generated, low-cycle fatigue cracks, the lengths of which have been determined by examination with an optical microscope. Information about crack depths is obtained from destructive examination of cracks which have been generated under similar stress conditions.

The Effective Reflectivity method [8] is another POD methodology that was also developed in the early 1980s that uses the properties of the normal distribution. This method was developed to address the challenges presented by the need to estimate POD for ultrasonic inspection of subsurface inclusions in forged metals. The lack of any method for generating artificial flaws that realistically simulated the properties of natural flaws was rendered inapplicable in earlier methods outlined above. A technique analogous to the laboratory-generated, low-cycle fatigue cracks that are believed to simulate natural low-cycle fatigue cracks is lacking. Instead, the method focused on the natural flaws that produced indications that were detected by the ultrasonic inspection. For each such indication, an estimate was made of the EFBH size. Some of these flaws were then examined metallographically, with unusually thin layers removed between each metallographic examination, in order to determine the three-dimensional size and shape of the inclusion. The EFBH size of each of the inclusions was studied and compared with the maximum cross-sectional area in a plane perpendicular to the detecting sound beam. This ratio, known as the Effective Reflectivity, R_e , (since it represented the difference between the response of the inclusion and a planar void) became part of a statistical database.

It has been found that either R_e or $\ln(R_e)$ approaches normally distributed data for all combinations of transducer, material, and flaw types that have been studied. The R_e method has been used to generate curves showing the dependence on flaw size of the mean POD and the associated lower one-sided 95% confidence limit. This was accomplished by combining this property with the assumption that ultrasonic response from inclusions is adequately described by

⁵ This model was selected based on empirical observation that it provided a better fit to experimental data than any of several competing models under review.

the same linear area-amplitude relationship [9] that is widely used to describe the response from FBHs. Due to the relatively rare occurrence of inclusions in modern aircraft engine forged materials, the R_e databases are, unavoidably, relatively small. One consequence of this is that straightforward linear regression techniques, such as are used in the \hat{a} versus a method, tend to produce physically implausible results. This effect is countered by reliance on the physical model to help determine the regression line. In other respects, it has many resemblances to \hat{a} versus a .

The Effective Reflectivity approach successfully circumvents the difficulty of generating synthetic flaws that accurately represent the properties of naturally occurring subsurface flaws, such as forging inclusions. However, the major disadvantage is that the process of characterizing the parameters of natural flaws, that are used in estimating the POD, renders them unusable for further POD studies. Thus, the redetermination of POD for each new circumstance can only be done by assembling new samples containing ultrasonic indications.

5.1.3 Methods Using Signal and Noise Distributions.

In all of the methods described above, background noise—from sources other than flaws—may appear to have been ignored. In fact, noise only enters into the decision-making process when setting a practical lower limit to the level of the decision threshold that distinguishes the detection from nondetection. The sources of this noise are usually known (for example, random electronic fluctuations in the instrumentation, ultrasonic reflections from grain-boundaries, or the effects of small instabilities in the position of the transducer with respect to the component undergoing inspection) but the distribution of the noise signals has rarely been studied. Instead, what are in effect the typical noise amplitudes in the upper tail of such a distribution known as a result of prescanning a specific component or experience accumulated during the earlier inspection of numerous similar components. This pseudopeak noise is used to set a practicable lower limit for the decision threshold, typically no lower than twice the peak noise, is a value that experience shows leads to an acceptably low level of false calls (i.e., occasions when the decision threshold is exceeded by a noise signal).

The full distribution of noise in the detection process has been known for many years, but until the ETC program, appears to have been ignored in NDE applications. Formal study of the theory of detection process started in the 1940s with the initial goal of optimizing the performance of radar [10] and sonar [11] detection systems. For example, the initial theoretical formulation of the radar detectability target, postulated normally, distributed noise which was fed to a linear detector that produced a Rayleigh distribution for the envelope of the detected amplitudes. Adding the target response signals to the noise was equivalent to increasing the scale parameter, σ , of the distribution. This is illustrated in figure 4, where, conceptually, the three curves may be thought of as distributions of (1) noise, (2) small-target signals plus noise, and (3) larger-target signals plus noise. For example, the POD for the small target is given by the proportion of the total area under the small-target curve that lies to the right of the decision threshold. The proportion of the total area under the noise curve to the right of the decision threshold is equal to the probability of false alarm (PFA).

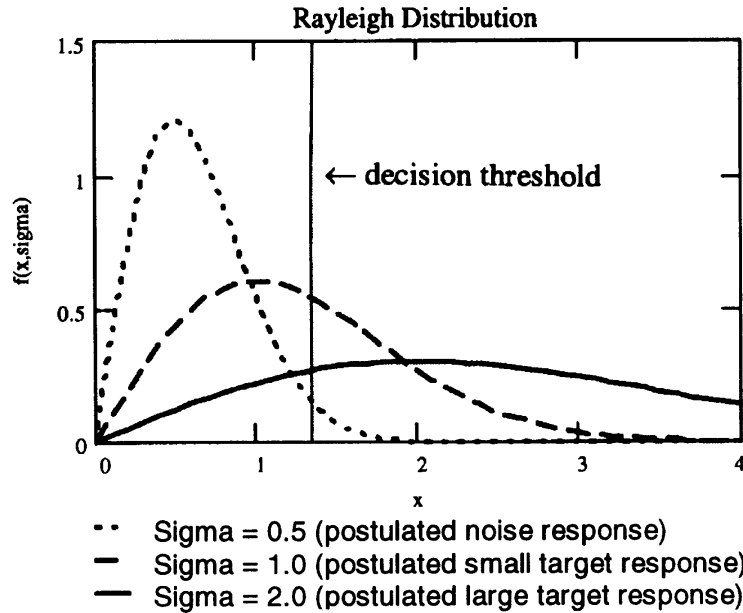


FIGURE 4. EXAMPLE OF RAYLEIGH DISTRIBUTIONS WITH INCREASING SCALE, SIGMA

The nature of the assumptions made about the distributions of signal and noise resulting from the early analyses of radar detectability may or may not be inappropriate to NDE processes. Figure 5 shows the consequences of a different set of assumptions, namely that the presence of a flaw modifies the location and not the scale of the distribution. The interpretation is similar in figures 4 and 5. It may be seen that (1) lowering the decision threshold will increase the POD but at the expense of an increase in the PFA and (2) large targets will usually have higher POD than small targets. The large targets clearly have a larger signal-to-noise (S/N) ratio than the small targets. Thus (once a suitable definition for it has been found⁶) the S/N ratio is a useful quantity by which to measure relative detectability.

Rice, one of the primary developers of statistical detection theory, developed the concepts (presented in figure 5) into families of curves by plotting POD versus the S/N ratio with PFA as the independent parameter. Peterson and Birdsall [12] subsequently plotted POD against PFA with S/N ratio as the independent parameter. This latter format has become known as the Relative (or Receiver) Operating Characteristic (ROC). It is often used as a convenient means for presenting empirical data in a wide range of fields (including cognitive discrimination between sensory stimuli) without considering the formulation of any underlying mathematical theory of detection. Occasionally, ROC has been used in presenting the results from studies of NDE, especially in the nuclear power industry [13]. It has also been found useful as a means for presenting POD and PFA data generated from mathematical models of inspection processes [14].

⁶ For figure 5, for example, a simple and useful definition is that the S/N ratio is equal to the ratio of the difference in means divided by the standard deviation (or the difference in location parameters divided by the scale).

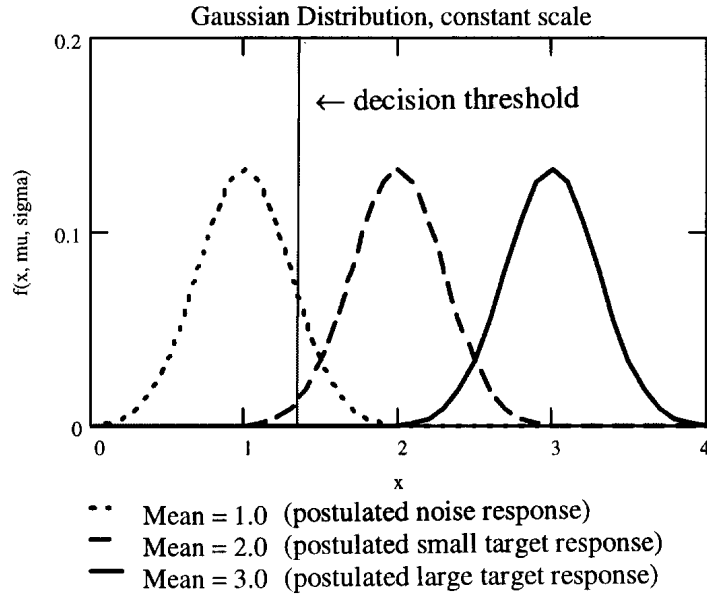


FIGURE 5. EXAMPLE OF GAUSSIAN DISTRIBUTIONS WITH INCREASING LOCATION (MEAN) AND CONSTANT SCALE (STANDARD DEVIATION)

5.1.4 Limitations of Existing Methods.

With the exception of the R_e method, all existing POD methods use sample flaws designed to simulate naturally occurring flaws; the existence of which is known prior to attempting the POD data acquisition, and the size of which may be determined either before or after the data acquisition. This represents a severe limitation for the application of these methods in the ultrasonic detection of subsurface flaws. When such naturally occurring flaws have been examined in detail, they typically have complex shapes and chemical compositions; therefore, it is extremely difficult to accurately simulate them. Acquiring this information about natural flaw morphology is difficult, time-consuming, and expensive. The few natural flaw types that have been investigated in sufficient detail may serve as a baseline for establishing natural flaw properties and their distribution within the material. The natural flaws have not been able to be characterized ultrasonically, or their accurate properties determined, until the sample is metallurgically sectioned. Because the sample has been destroyed, it is not possible to use it again and perform an ultrasonic evaluation.

Even if the natural flaw properties were known and an adequate simulation could be accomplished, the manner in which ultrasonic response depends on the flaw and material properties would necessitate locating such simulated flaws, representing a wide variety of flaw properties, at a wide variety of physical locations.

Although the R_e method avoids the need to make simulated flaws and intrinsically guarantees that POD data acquisition is based on the true properties of natural flaws, it is limited by the small number of detectable flaws. Recent modifications to the method have included adding a provision of a statistically based means for compensating the hypothesized missing data.

However, the detailed metallographic examination of several thousand pounds of alloy would be necessary to determine if flaws actually escaped detection. This is totally impracticable. Furthermore, while the linear area-amplitude response embodied in the simple FBH model is plausible for the small flaws found in forgings (and especially in powder metals), it is uncertain whether it can be applied to much larger flaws that may occur in billets. Therefore, a more satisfactory model is clearly needed.

All the methods discussed share the major limitation that POD estimates strictly apply only to the circumstances under which the POD measurements took place. This results from the current lack of methods for accurately predicting the effects on POD of changing any one (or more) of the numerous parameters that can affect detectability of flaws. These include flaw, material, inspection equipment, calibration, and scanning parameters (and, at least for nonautomated inspections, human factors too). This limitation means that every time one of these parameters is changed (such as a different type of instrument or transducer or a change in the scan increment for example) a new determination of POD is necessary. While it has proved possible to do this in estimating POD for eddy-current inspection of low-cycle fatigue cracks, it is no simple matter for the R_e method because the original sample defects would have been examined destructively in order to determine their true size and are no longer available for repeated use.

It thus appears that several changes in POD methodology are highly desirable; in particular:

- Necessary mathematical models should be developed (and then validated by experiment), that (1) would model the response of flaws that are comparable to or larger than the sound beam and (2) would provide a basis for predicting the effects of changing as many of the other inspection parameters as possible and
- A means for taking PFA into account should be introduced while still maintaining statistical rigor.

A combination of advanced modeling techniques with the S/N-based detection theory concepts originally developed for radar applications should provide a basis for producing accurate POD estimates over a range of conditions prevalent in practice.

5.2 ETC IMPLEMENTATION.

5.2.1 Background.

As noted at the end of section 5.1.4, several changes need to be made to the POD methodology (while maintaining statistical rigor), i.e., a means for taking PFA into account and the introduction of the sophisticated models that describe ultrasonic signals. In this section, the approach that was developed by the ETC is described.

However, before entering into a detailed discussion, some heuristics may be helpful in setting the stage. A central difficulty in estimating the POD of naturally occurring flaws using existing empirical techniques is their complex shape and chemical compositions which render them very difficult to simulate in the laboratory, both because of lack of information about their natural morphology and the difficulty in producing laboratory samples which mimic that morphology. This difficulty is exacerbated by the extremely low frequency of occurrence of hard-alpha inclusions in the field, and the fact that these are typically destroyed in the process of identifying their origin.

The general motivation for introducing advanced modeling techniques has been given in section 5.1.4. In the context of the small number of naturally occurring hard-alpha inclusions that are found, it can be put in another way, namely, that the physical understanding incorporated in a model allows one to extract as much information as accurately as possible from the available data. A simple example illustrates the point, as shown in figure 6. The first step in methods based on the full-flaw response (section 5.1.2) involves determining the distribution relating flaw response to flaw size. Suppose one has a set of flaw response data which one wishes to relate to flaw size by a regression line and, as shown conceptually in figure 6(a), there is considerable scatter in the data. Standard linear regression yields the fit shown, but it can be seen intuitively that there is considerable uncertainty in the slope of this regression result. However, if one had further information such that the flaw response vanishes for zero flaw size and is proportional to flaw area, considerable improvements can be made, as shown in figure 6(b). Taking size to be area and forcing the regression through the origin leads to a substantial reduction in uncertainty. In essence, through the use of this elementary model, one is able to require the data to define only a slope rather than a slope and an intercept⁷. This idea is implicit in the R_e approach. The methodology developed in ETC Phase I, and to be described later in this report, represents a higher level of sophistication through the use of more accurate flaw response models.

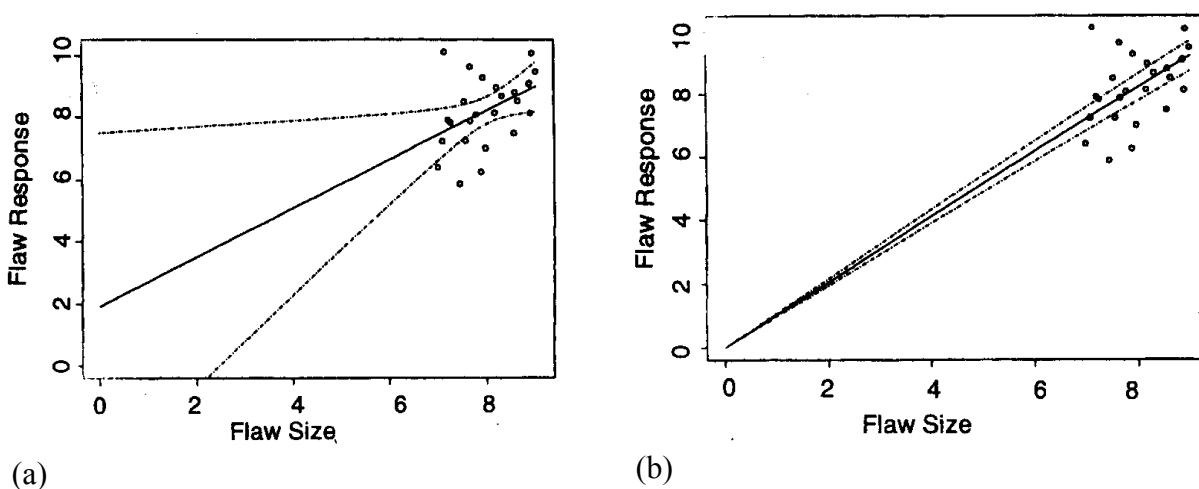


FIGURE 6. INFLUENCE OF ADDITIONAL INFORMATION ON ACCURACY OF A LINEAR REGRESSION FIT TO DATA

⁷ When \hat{a} is plotted versus a in logarithmic space, the analogous constraint is that the plot has a slope of unity.

The solid line represents the fit determined by regression, and the dashed lines denote the uncertainty in that fit.

- a. Standard regression has large uncertainty in slope.
- b. Model-derived knowledge that the line must pass through the origin substantially reduces the uncertainties in the slope.

Some important distinctions can be made by using a physical model as part of the determination of POD which cannot be addressed using traditional approaches. In an empirical test, when an indication is observed, there is no way to know whether it is because the signal from a flaw exceeded the threshold or if the indication is a consequence of noise or some other false cause. This uncertainty exists even when the probe is positioned so that a flaw is in the field of view. However, by using the information from physical models of the flaw response, it possible to directly consider this distinction.

To facilitate the discussion of these two cases, the term probability of true detection (POTD) is introduced to indicate the situation when an indication is truly due to the signal from a flaw. The term probability of an indication (POI) is used to indicate the situation when an indication is due to any cause. As figure 7 shows, one would expect the POTD $\rightarrow 0$ and the POI \rightarrow PFA as the flaw size vanishes. The difference between the POTD and POI is, of course, the probability that a signal is a reflection from a benign inhomogeneity such as a grain boundary.

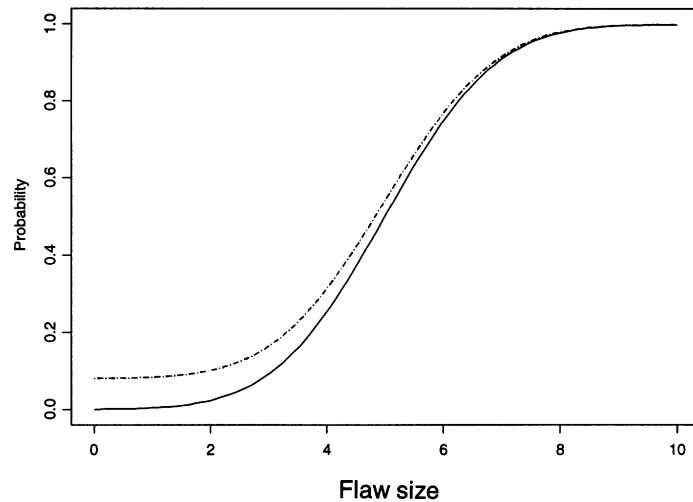


FIGURE 7. A SCHEMATIC VIEW OF THE RELATIONSHIP BETWEEN POI AND POTD
(The solid line represents POTD and the dashed line represents POI.)

These concepts can be further clarified by making reference to the hypothetical distributions of flaw response plus noise that were presented in figures 4 and 5. In the case illustrated by figure 4, the response-plus-noise distribution is broadened and shifted by increases in the signal, whereas it is only shifted in the case illustrated in figure 5. For either of these cases, the fraction of the area to the right of a specified threshold (decision threshold line shown in figures 4 and 5)

is considered the POI (since one does not know whether a response above threshold is due to the noise or the flaw response). As flaw size approaches zero, the value that POI approaches will be determined by how rapidly the probability density function for the noise response drops to zero in the large response tail of the distribution. For the case in which the distribution shifts without changing scale (figure 5), this limit can become quite small, producing a POI curve which qualitatively looks like the lower (POTD) curve in figure 7.

There are alternative definitions of false alarms. In this work, one used the definition that a false alarm occurs when the threshold is exceeded and no flaw is present. An alternative view would say that a false alarm occurs when an acceptable size flaw will return an unacceptable signal (i.e., a signal exceeding threshold). An illustration is found in figure 4, if one considers the small target to be acceptable and the large target to be unacceptable.

This definition of false alarm plays a role in typical fracture mechanics applications. For example, an engineer needs to define a flaw size that must be detected with 90% probability in order for a component, hypothetically containing such a flaw, to survive a given number of stress cycles without catastrophic failure. An inspection threshold is selected corresponding to 90% POD for that size flaw and 10% of such flaws will likely be missed. In addition, a significant proportion of smaller flaws (smaller than the critical size) will give signals that are larger than the inspection threshold. In this situation, two approaches to the lifing calculation present themselves. The engineer may make use of the entire curve of POD versus flaw size or choose to replace the POD curve by a step function at the critical flaw size, with POD going from zero for smaller flaws to 1 for larger flaws. In the former approach, large signals from small flaws are just part of the total picture. However, it appears that in the latter approach, an above-threshold signal from a below critical size flaw is truly a false call.

This definition of a false call, as a unacceptable signal from an acceptable flaw, will not be considered in this work. However, it should be emphasized that it could be readily considered when using basic tools, i.e., distributions of noise and signal plus noise, which were developed using the new methodology. Indeed, it is the ability to address such questions as the distinction between POTD and POI and the definition of PFA that is one of the major advantages of the new methodology.

Because existing tools have not been available to consider such issues, their distinctions have not been fully incorporated in the analyses of the POD by the life management community. For example, Spencer used the term POD to include cases where the threshold is exceeded for reasons independent of crack lengths, i.e., false calls [15 and 16]. It is the impression of some members of the team preparing this report that POTD of the greatest interest to the life management community, but that POI may become so in the future. A dialogue between these two communities is required to address such issues. Until such time, the goal is to determine POTD.

5.2.2 Distributions-of-Noise and Flaw-Plus-Noise Signals.

For the NDE techniques that generate background signals with randomly varying amplitudes (e.g., ultrasonics, eddy-current, nonfilm radiography, etc.), there will be a distribution of signals that would be observed in the absence of a flaw (noise distribution). The distributions of flaw signals that would be observed when a flaw is present and its response is modified by the noise (flaw-plus-noise distribution). Conceptually, one can define the flaw-plus-noise distribution that corresponds to the distinction between a true defect or signal indication of a defect. In the development of the methodology, the aim was to determine the distribution of true flaw signals in the presence of noise, which will be employed in the majority of this section. The alternate definition, the distribution of indications, will be considered in section 5.2.11. The distribution of flaw and noise signals would depend on the actual character of the flaw and its position in the material background. Because of background noise, there is a chance that a flaw will not be detected (a miss) or that a noise signal could be mistaken for a flaw signal (a false alarm).

Figure 8 shows examples of a possible distribution of noise signals and several distributions of flaw and noise signals. POD and PFA are equal to the areas indicated by the shading. A flaw-plus-noise distribution describes the distribution of signals that one would see from a population of flaws having a particular set of describable characteristics, e.g., a size, measure. When the set of characteristics changes, the distribution changes. If circumstances allow, these distributions might be estimated empirically. This might be practicable for surface-connected flaws (such as low-cycle fatigue cracks) where there may be an abundance of data relating observed signal-to-flaw characteristics. For other situations where it is difficult or impossible to obtain enough data to adequately characterize the observed signal's relationship to flaw characteristics, physical/mathematical models (properly verified for the needed application) can be used with the limited available data to provide estimates of inspection capability. Ultrasonic inspection for detection of subsurface flaws falls into this latter category.

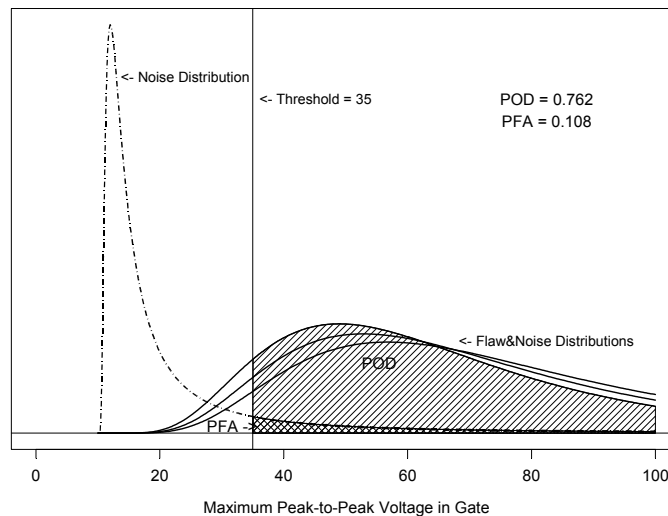


FIGURE 8. EXAMPLES OF POSSIBLE DISTRIBUTION OF NOISE SIGNALS AND DISTRIBUTIONS OF FLAW-PLUS-NOISE SIGNALS, SHOWING POD AND PFA BY THE CROSS-HATCHED AREAS

5.2.3 General Approach.

The general approach for assessing POD in this program was based upon the following steps:

- A physical model, based on the theory of ultrasonic wave scattering, provides predictions for typical or expected response measurements for a given set of conditions.
- Experimental data are used to compare the predictions of the physical model with actual recorded signals to determine the distribution of the deviations or residuals, which are represented by the symbol $\hat{\epsilon}$. A statistical model is used to quantify “deviations” between the physical model predictions and actual NDE measurements. This statistical model for the deviations describes the sources of variability in ultrasonic signals that are not included in the physical model. This model also provides a framework for predicting PFA and POD.
- POD can be assessed for some specified ranges of inspection conditions, different materials, and different defect types. Limitations of the approach are due to the breadth and adequacy of the physical and statistical models in which there are deviations between signal and the predictions made by the physical model.

The steps in implementing the methodology are presented in more detail in figure 9. Required input data are shown in the ellipses at the top, while input parameters defining the inspection problem are shown in the ellipses at the left. A beam-centered flaw (BCF) provides a model for determining NDE capability, the POD for a flaw centered in the focal region of the beam (as would occur after peaking the signal) or a scan plan having very close-spaced positions between adjacent transducer pulses. This statistical model is developed to describe the deviations between the (microstructure free) model predictions and the experimental data. A similar approach is employed for coarser inspection-scan plans. This suggests the possibility of the flaw being substantially off the beam center. This report will present the methodology and discuss its application to FBHs and SHAs. Application to naturally occurring flaw data awaits the conclusion of the Contaminated Billet Study (CBS).

The general approach for assessing PFAs is based on the empirical determination of noise signals distribution from scanned data taken on a block for probe positions. This is done far away from flaws, so there would be no effect from the flaw in the measurements.

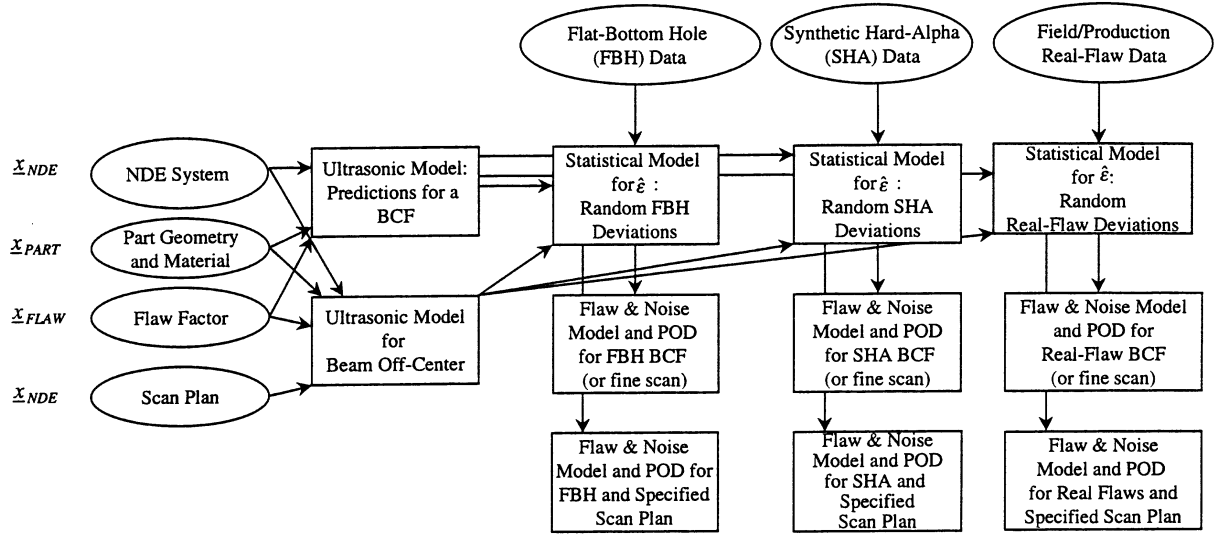


FIGURE 9. PROPOSED ETC POD ESTIMATION METHODOLOGY

Details of the determination of POD and PFA are presented in the following sections.

5.2.4 Fixed and Random Factors Affecting Flaw and Noise Distributions.

Conditional on a set of specified fixed factors that affect the signal strength, the cumulative probability distribution for flaw and noise signal can be expressed as

$$\Pr(Y \leq y) = F(y; \underline{x}, \underline{\theta}) \quad (3)$$

with a corresponding probability density function $f(y; \underline{x}, \underline{\theta}) = dF(y; \underline{x}, \underline{\theta}) / dy$. This is the flaw and noise model identified in several boxes in figure 9.

Here $\underline{\theta}$ is a vector of parameters that is, for the most part, independent of \underline{x} , where $\underline{x} = (x_{FLAW}, x_{NDE}, x_{PART})$ is a vector of factors that affect the ultrasonic signal response. In particular,

- x_{NDE} contains NDE system factors like transducer, scan plan, and electronic system characteristics.
- x_{PART} contains PART factors like part geometry, type of material being inspected, surface roughness, etc.
- x_{FLAW} contains FLAW factors like size, density, shape, composition, degree of voiding/cracking, and orientation/position relative to the ultrasonic beam.

The ETC model for noise-only signals (UT signal when there is no flaw illuminated by the beam) is similar, except that the distribution would not depend on \underline{x}_{FLAW} . To partition the factors into fixed factors and random factors, a signal-determining factor is taken to be fixed if either

- the factor can be controlled in the inspection operation (e.g., frequency, transducer parameters, scan increment, pulse rate, gate width, etc.) or
- it is desired to estimate POD as a function of the factor level (e.g., flaw size and depth).

There are a number of other factors determining flaw-plus-noise signals that will be considered to be random during production/field inspection. These include:

- Details of microstructure, including position of a flaw relative to grain boundaries in material (material effects).
- Flaw position relative to the ultrasonic transducer.
- Flaw morphology, including shape, orientation, composition, and extent of voiding (e.g., real flaws tend to have complicated shapes).

Referring again to figure 9, the strategy is to build the controlled inspection parameters into the ultrasonic model, that will

- infer the effects of microstructure from the responses of nominally identical FBHs and SHAs.
- treat variations in flaw position with the ultrasonic model.
- infer the flaw morphology effects from experiments on naturally occurring defects.

5.2.5 Measurements of Responses of Nominally Identical Flaws.

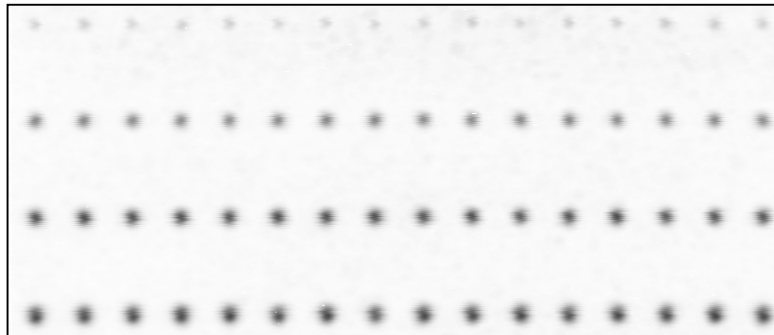
Figure 9 shows that the measurement of the response of simulated or real flaws serves as the experimental input for the methodology. In the case of FBHs and SHAs, the experimental input consists of the response measurement for a set of nominally identical scatters. In all cases, the measurement corresponds to the peak-to-peak signal excursion within a time gate. Testing the validity of that model can be done by comparing the means of this response to the ultrasonic model. The deviations describe the material contributions of the variability flaw response. The ultrasonic models and the details of these experiments are described in section 6 and appendix A. The following paragraphs identify those results which form the basis of the illustrative calculations.

A factorial experiment was first conducted to obtain information on the distribution of flaw-plus-noise signals for FBHs in titanium [17]. The experiments used for the computations described in

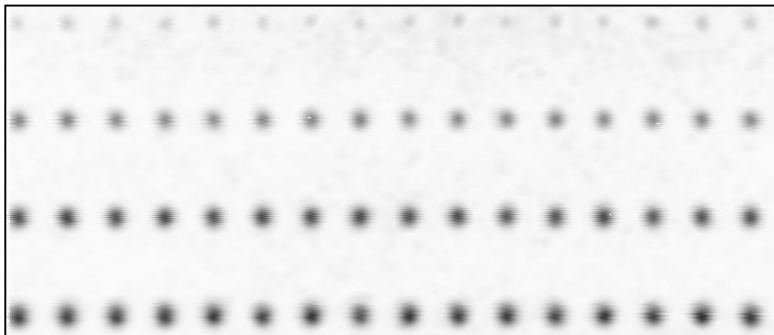
this report were conducted using 5-MHz transducers focused at depths of 0.5", 1", and 1.25" and incident angles of -2.5°, 0°, and 5° with scan increments of 0.010" in both the x and y directions. The sample was fabricated from a Ti-6Al-4V ring forging, machined into a flat plate and contained 64 FBHs (16 each of sizes nos. 1, 2, 3, and 4) with bottoms 1" below the inspection surface. Figure 10 presents C-scans for three of the cases. Voltage readings were taken from each of the 16 nominally identical nos. 1, 2, 3, and 4 FBHs. Full details are provided in appendix A.



FBH C-scan: 5-MHz transducer no. 2, normal incidence, focused at 1" depth on FBHs



FBH C-scan: 5-MHz transducer no. 2, 5 deg. tilt in water, focused at 1" depth on FBHs



FBH C-scan: 5-MHz transducer no. 2, 5 deg. tilt in water, focused at 0.5" depth above FBHs

FIGURE 10. C-SCANS FOR THREE OF THE CASES CONSIDERED IN THE FBH FACTORIAL EXPERIMENT

A similar factorial experiment was conducted to obtain information on the distribution of flaw-plus-noise signals for synthetic hard-alpha inclusions (5.9% nitrogen by volume) in a Ti-6Al-4V alloy [18 and 19]. Voltage readings were taken on each of eight nominally identical nos. 2, 3, 4, and 5 cylindrical synthetic hard-alpha inclusions. These were vertically oriented (i.e., the major axes were perpendicular to the inspection surface) with the nearest of their circular end faces 1" below the inspection surface. The synthetic hard-alpha inclusions were produced by procedures developed by Gigliotti, et al. [2] using hot isostatic pressing (HIP) consolidation procedures to blend a mix of Ti and TiN powder. These inclusions were embedded in material taken from a Ti-6Al-4V ring forging via HIP bonding procedures. Cylindrically shaped inclusions were inserted in matching holes drilled on the face of the Ti-6Al-4V block. The sample was then covered with a matching piece of Ti-6Al-4V, the edges were welded together in vacuum, and the cover plate was HIP-bonded to the block. The HIP bonding conditions were selected to assure good bonding, to minimize nitrogen diffusion from the inclusion, and to minimize changes in the phase volume and microstructure of the Ti-6Al-4V alloy. Here the flaw size measure was adapted from the FBH convention where no. 2 is 2/64 in., no. 3 is 3/64 in., etc. The part of the experiment used for the computations in this report was conducted using the same Panametrics 5-MHz-focused transducers used in the FBH experiment as well as a special 10-MHz-focused transducer. Data were taken at focal depths of 0.5, 1, and 1.25 inches and incident angles of 0°, 2.5°, and 5° with scan increments of 0.005" in both the x and y directions. Full details are also provided in appendix A.

Figure 11 provides an example of one of the C-scans for the SHA case. From this, graphs of flaw response versus position were extracted and then compared to the predictions of the physical model. Figure 12 shows a graph of a subset of that data. Each frame shows signal strength (mV) for each of the eight nominally identical no. 5 SHA flaws as a function of the x-dimension offset distance from the center of the flaw. The seven frames in the vertical column show strengths as a function of the y-dimension offset. Each column is for a different focal depth, as indicated. The bold line shows the corresponding physical model predictions.



5.9% N SHA C-scan(a): 10-MHz transducer no. 4, normal incidence, focused at 1" depth on SHAs

FIGURE 11. C-SCAN OF SHA BLOCK WITH 10-MHz PROBE FOCUSED IN THE PLANE OF THE FLAWS AND ILLUMINATING THE BLOCK AT NORMAL INCIDENCE

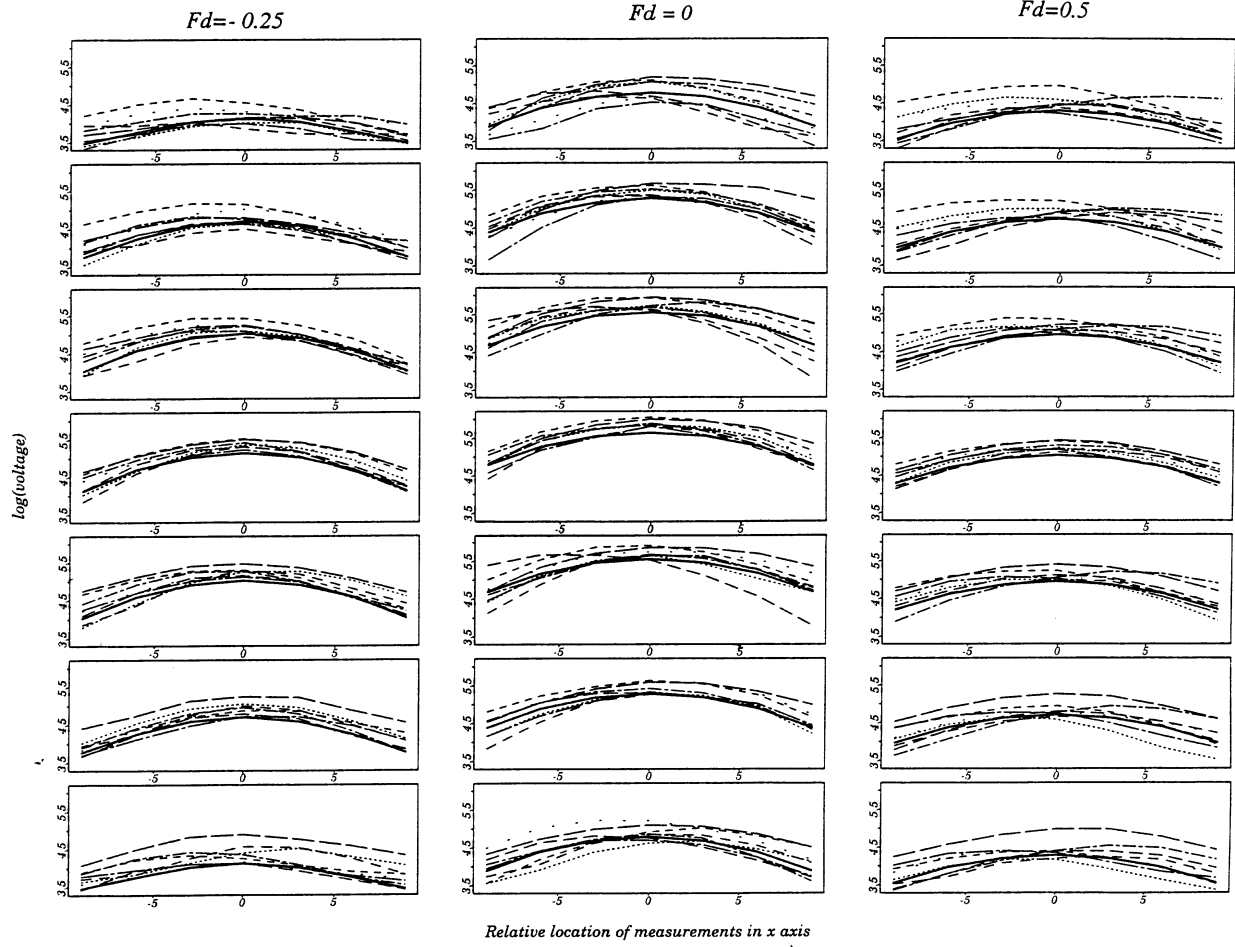


FIGURE 12. PLOT OF DATA FROM THE 10-MHz-FOCUSED TRANSDUCER, ILLUMINATING THE 5.9% N SHA BLOCK AT NORMAL INCIDENCE

It is through the analysis of the deviations between that data and model predictions that yields the statistical model for $\hat{\epsilon}$, the random SHA deviations, that was shown in figure 9.

The plots shown are for the no. 5 cylindrical SHAs. As described in the text, the columns correspond to different positions of the focal plane with respect to the depth of the end of the SHA nearest to the inspection surface. The rows correspond to different offsets of the scan lines from the SHA centers.

5.2.6 Modeling of the Generalized Signal/Prediction Deviations.

The ultrasonic NDE model (UNDE model) will predict the flaw signal as a function of the fixed factors in \underline{x}_{FLAW} , \underline{x}_{NDE} , and \underline{x}_{PART} . The physical model used here, which neglects any contributions of microstructural effects to the flaw signal, will be described in section 6 and appendix A [17, 18, and 20]. In this first implementation of the methodology, a statistical model

was used to describe deviations of the ultrasonic response from the physical model predictions. These variations were attributed to random factors related to microstructure, flaw fabrication measurement error, and possibly model errors.

By examining the relative contributions of these potential sources of variability, the effects of measurement error on laboratory measurements can be controlled. This variation can be reduced by carefully selecting parameters (e.g., sufficiently small scan indices and gate widths) or by accounting, in model predictions, for the off-center distance (note the beam off-center case in figure 9). The degree of uncertainty associated with flaw fabrication variability is unknown at the present time. Progress towards extending the physical models to predict microstructure-controlled distributions of signals, which will reduce the need for empirical experiments, is discussed in section 9.4. The degree to which the empirical variabilities observed in these laboratory tests can be generalized to full-scale components depends on the relative contributions of these effects. This is planned to be examined further in future work.

Y denotes the experimental voltage and \tilde{Y} ⁸ denotes the UNDE model prediction for Y . The prediction \tilde{Y} is a function of \underline{x}_{FLAW} , \underline{x}_{NDE} , and \underline{x}_{PART} and provides a prediction which is initially assumed to be at the center (location) of the flaw-plus-noise signal distribution. Note this assumption is not valid when the flaw signal approaches the noise level, which will be discussed in section 9.4. The authors found it convenient to describe generalized deviations between the UNDE predictions and the actual data, using Box-Cox transformations, as

$$Deviation = g(Y; \lambda, x) = \begin{cases} \frac{(Y)^\lambda - 1}{\lambda} - \frac{(\tilde{Y})^\lambda - 1}{\lambda}, \lambda \neq 0 \\ \log(Y) - \log(\tilde{Y}), \lambda = 0 \end{cases} \quad (4)$$

Here λ is a constant, known as the Box-Cox transformation parameter. It is determined empirically from the data, where the aim is to select a value such that the transformed data follows a standard distribution, e.g., the normal distribution, with standard deviation independent of flaw size. These generalized deviations provide the shape and spread of the flaw-plus-noise signal distribution. The value of λ was also chosen empirically to equalize variance (with respect to flaw size) and otherwise make distributions, as much as possible, independent of the factors $\underline{x} = (\underline{x}_{FLAW}, \underline{x}_{NDE}, \underline{x}_{PART})$ that drive the UNDE model. In the special case of $\lambda=1$, the deviations would be purely additive to the model prediction.

This approach was applied to both the FBH and SHA data. In each case, the shape of the distribution of the deviations was investigated for different values of the transformation parameter λ . This same value of $\lambda=0.3$ was reported

- by Meeker, et al. [20] as appropriate for stabilizing the distribution of generalized deviations from UT signals from flat-bottom holes,

⁸ Read as “Y hook” or, more formally, as “Y breve.”

- by Meeker, et al. [21] as appropriate for stabilizing the distribution of generalized deviations from UT signals from synthetic hard-alpha inclusions, and
- by Sarkar, et al. [22] for stabilizing the distribution of generalized deviations from UT signals from cracks in nuclear power plant steam generator tubes.

Although these empirical studies do not mandate that $\lambda=0.3$ must be used, they do suggest that the result has significant generality. Figure 13 shows normal, logistic, and largest extreme value (LEV) distribution probability plots for the deviations from the SHA case when the ultrasonic beam ensonified the flaw at normal incidence. This figure indicates that normal and logistic distributions provide adequate distributional fits and that the variance of the deviations is approximately equal over the different flaw sizes. Deviations for the no. 2-sized inclusions were omitted from this analysis because the UNDE model predictions for flaw responses were far below the background noise level. Given these results, the normal distribution was used to describe the distribution of generalized deviations.

The omission of the no. 2 FBH data from the boxplot in figure 13, because the UNDE model predictions were far below the background noise, assumes that the distribution so determined is for true detects. In the limit when no flaw is present, equation 4 approaches the limit Y^λ / λ , or $Y^{0.3} / 0.3$. The implications of this on the prediction of POTD will be discussed in the next subsection.

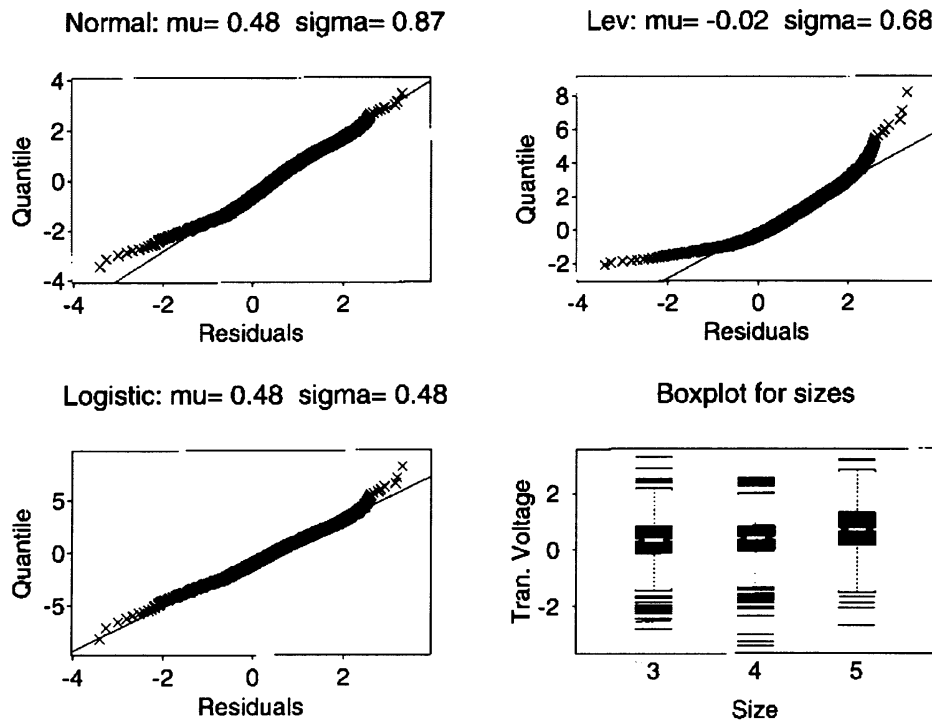


FIGURE 13. DISTRIBUTION OF BEAM OFF-CENTER DEVIATIONS AFTER $\lambda=0.3$ TRANSFORMATION FROM THE 10-MHz-FOCUSED TRANSDUCER, NORMAL INCIDENCE SCANS OF CYLINDRICAL, SYNTHETIC HARD-ALPHA INCLUSIONS

5.2.7 Basic Probability of a True Detection (POTD).

An ultrasonic indication is said to occur when $Y > y_{thresh}$, where Y is a measured gated peak-to-peak amplitude and y_{thresh} can be set according to specified user criteria (e.g., to make the probability of a false alarm essentially 0 or to minimize expected risk). This event is called a detection if a flaw is present.

For some applications, it may be of interest to compute POTD values for one or more sets of fixed values of all of the components in $\underline{x} = (x_{FLAW}, x_{NDE}, x_{PART})$. It is called the basic POTD. This corresponds to the probability of obtaining a flaw signal above the detection threshold when the transducer has a fixed position with respect to a flaw that has specified characteristics. As will be discussed in section 5.2.8, the actual POTD observed in a production inspection will be determined by integrating a number of these factors corresponding to the quantities which are unknown in the inspection and hence treated as random variables. Based on the Box-Cox transformation following the general model presented in section 5.2.6, the deviations are normally distributed. Then

$$\Pr(Voltage < Threshold\ Voltage) = \Pr(Y \leq y_{thresh}) =$$

$$\Pr[g(Y) \leq g(y_{thresh})] = \Phi \left[\frac{g(y_{thresh}) - \mu_g}{\sigma_g} \right] \quad (5)$$

where Φ is the standard normal (Gaussian) cumulative distribution function and μ_g and σ_g are computed from the available deviation data. Then the possibility of a detection on any given reading is

$$POTD(x) = \Pr(Y > y_{thresh} | \underline{x}) = 1 - \Phi \left[\frac{g(y_{thresh}) - \mu_g}{\sigma_g} \right] \quad (6)$$

As in Meeker, et al. [20], this is called the Basic POTD. Note that the function $g()$ in the calculation in equation 6 depends on \underline{x} through \tilde{Y} , the UNDE prediction for Y , is defined in equation 4. The parameter λ in equation 4 was chosen to make the spread in the generalized deviations approximately constant across all values of \underline{x} . The parameter σ_g is a measure of spread in the generalized deviations. Because the noise is generated by the material, it is reasonable to assume, for a material specimen with a given noise level, that σ_g would be constant at other \underline{x} values within the range of the data (depth, flaw size, tilt angle) and for other transducers that can be modeled adequately. Relatedly, μ_g measures systematic model bias (i.e., tendency to be off in one direction or the other over different). If the UNDE model provides adequate unbiased predictions, then one could set $\mu_g = 0$. In certain circumstances, however, a bias correction would be warranted.

It is useful to study the behavior of POTD (\underline{x}) as $\underline{x} \rightarrow \underline{x}_{no\ flaw}$. For simplicity, suppose that $\mu_g = 0$ (no model bias). Then for fixed y_{thresh} , the limit of the generalized deviation $g(y_{thresh})$ as $\underline{x} \rightarrow \underline{x}_{no\ flaw}$ is $y_{thresh}^\lambda / \lambda$ for $\lambda > 0$. Operationally, if σ_g is small (e.g., small enough so that $(y_{thresh}^\lambda / \lambda) / \sigma_g > 4$ then $POTD(\underline{x}) \rightarrow 0$ (approximately) as $\underline{x} \rightarrow \underline{x}_{no\ flaw}$. This would imply that, with a low level of (empirically measured) variability, any signals would be well below the threshold for small or nonexistent flaws. If, on the other hand, σ_g is larger relative to $y_{thresh}^\lambda / \lambda$, $POTD(\underline{x}) \rightarrow 1 - \phi((y_{thresh}^\lambda / \lambda) / \sigma_g)$ as $\underline{x} \rightarrow \underline{x}_{no\ flaw}$. For example, if $(y_{thresh}^\lambda / \lambda) / \sigma_g = 1.645$ then $POTD(\underline{x}) \rightarrow 0.05$ as $\underline{x} \rightarrow \underline{x}_{no\ flaw}$. That is, if there is a lot of noise (being reflected in a large σ_g) then the model could predict an importantly large probability of a detect even if there is no flaw.

It should be noted that an implicit assumption in this discussion is that the distribution fitted to the generalized deviations, when the signal is greater than the noise, applies to “true defects” when the signal drops below the noise level.

5.2.8 POTD for Production Inspection.

To predict POTD for production inspection, it will be necessary to account for random factors in the inspection process such as flaw position relative to the beam. In an automatic system with gated detection, a volume of the material is interrogated with each pulse. A flaw at any position within that volume can be detected, yet the strength of the signal will depend on the flaw’s position within the volume, being roughly proportional to the square of the local strength of the beam. Thus, the longer the gate or the greater the scan increment, the broader the signal distribution due to the variation of ensonifying intensity. This effect is taken into account by integrating over possible flaw positions. This evaluation will require the evaluation of a joint distribution of random factors, this distribution is defined by the inspection process and system.

To illustrate this, the ETC will show how to evaluate the effect on POTD by using different scan increments. To keep the example simple, assume that the cylindrical synthetic hard-alpha flaw is, as in the experiment, vertically oriented (i.e., is oriented with its major axis perpendicular to the inspection surface) with the nearest of its circular end-faces 1” below the surface and that the beam is focused with normal incidence at that depth. To get POTD as a function of size a and scan increment, $\underline{x} = (a, \underline{x}_{FIXED}, \underline{x}_{RAN})$ was redefined, where \underline{x}_{RAN} is the two-dimensional position of the flaw in the block and \underline{x}_{FIXED} is a vector of all of the other factors in \underline{x} , which are assumed to be fixed. To compute POTD for fixed values of size a and \underline{x}_{FIXED} , equation 6 was integrated with respect to \underline{x}_{RAN} over the entire range of \underline{x}_{RAN} .

$$POTD(a, \underline{x}_{FIXED}) = \int f_{\underline{x}_{RAN}}(\underline{x}_{RAN}) POTD(a, \underline{x}_{FIXED}, \underline{x}_{RAN}) d\underline{x}_{RAN} \quad (7)$$

where $f_{\underline{x}_{RAN}}$ is the probability density function of \underline{x}_{RAN} .

For the SHA experiment, with random x and y flaw position in the plane and fixed focal depth, it was assumed that flaw position is uniformly distributed between scan lines as shown in figure 14. In other words, it was assumed that the origin of the scan grid is positioned randomly with respect to the flaw location, with no preferred position. Then all possible offsets of the flaw with respect to a scan grid can be considered by assuming the flaw to be uniformly distributed in the $1/4$ square shown in figure 14. Because of symmetry and similarity, the POTD, in this case, is easy to compute by simply integrating over the $1/4$ square shown in figure 14. If the signal response pattern was nonsymmetric, one would need to integrate over one of the larger squares. Then

$$POTD(a, \underline{x}_{FIXED}) = \iint f(x, y) POTD(a, \underline{x}_{FIXED} | x, y) dx dy \quad (8)$$

where $f(x, y)$ is a joint probability density function describing the x - y position of a flaw relative to the scan lines. Physically, a uniform distribution should provide an adequate description for this distribution for the reasons given above.

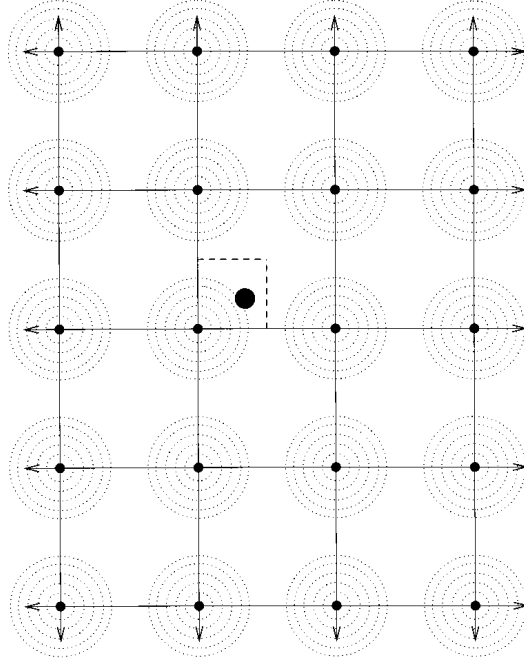


FIGURE 14. SCHEMATIC UT SCAN PLAN WITH COARSE INCREMENTS

This idea can easily be extended to randomness in the depth of the flaw (so focal depth, now denoted by z , is removed from \underline{x}_{FIXED} and used as a variable of integration). Finding POTD in this case will require integration over a joint distribution of x - y flaw position relative to scan lines and flaw depth (z) relative to focal depth. In particular,

$$POTD(a, \underline{x}_{FIXED}) = \iiint f(x, y, z) POTD(a, \underline{x}_{FIXED} | x, y, z) dx dy dz \quad (9)$$

Here $f(x, y, z)$ is a joint probability density function describing the x-y position of a flaw relative to the scan lines and the depth (z) relative to the focus depth. The x-y part of the distribution could be expected to be uniform as discussed above. The range of the z-part of the distribution, also expected to be uniform, would correspond to the depth range from which signals are accepted in the gate. In some situations, however, a nonuniform distribution could be expected in the depth dimension, depending on the distribution of flaws in material being inspected (e.g., there might be a higher probability to have flaws on or near the surface or in some other region, as determined by the processing history).

5.2.9 Probability of a False Alarm (PFA).

The probability of a false alarm can be defined as the probability of an above-threshold reading when there is no flaw. Under our model, the probability of such a false alarm on any given reading is an analogy to equation 6, is

$$PFA = \Pr(Y > y_{thresh} | \text{no flaw}, \underline{x}) = 1 - \Phi \left[\frac{g_n(y_{thresh}) - \mu_{gn}}{\sigma_{gn}} \right] \quad (10)$$

where g_n is the distribution of noise. In this work, based on empirical data (presented in section 9.2), g_n has been taken to be a lognormal transformation that can be used to represent the noise in the sample of interest. Section 9.3 discusses more sophisticated approaches that should be examined in future programs. The statistical model and parameters μ_{gn} and σ_{gn} are the means and standard deviations of the lognormal distribution representing the noise but using an average of the voltage signals over several different regions that do not contain flaws. Note that the statistical model for the noise is the limit $\lambda=0$ of the Box-Cox transformation.

5.2.10 Relative Operating Characteristics (ROC).

As noted in section 5.1.3, ROC curves are a simple way to display POTD and PFA information simultaneously and are a means to compare different inspection methods/conditions without having to specify a threshold. To compute an ROC curve, choose a fixed flaw size a , Vary y_{thresh} , compute PFA from equation 10, POTD from equation 7, and plot the resulting set of points as a curve on a graph of POTD versus PFA. Repeat this process for different values of size a , generating separate ROC curves for each value of a .

Figure 15 illustrates ROC curves and their utility. For a very low threshold, one expects high values for both POTD and PFA and, as the threshold is increased, both POTD and PFA will decrease. A good NDE technique will be differentiated from a poor NDE technique by using a form of this function. PFA could be substantially lowered before POTD was significantly affected. On the other hand, if the test were essentially a guess, both would react in the same way.

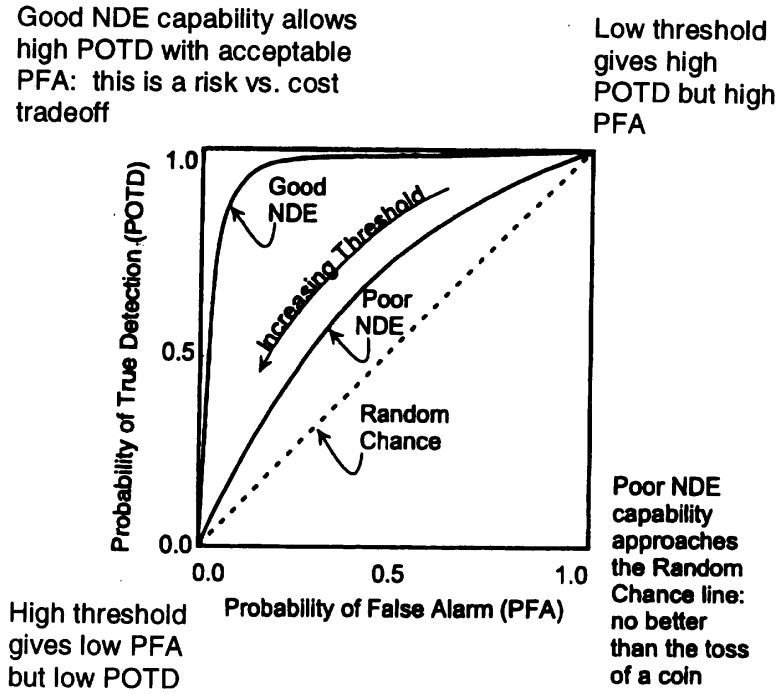


FIGURE 15. RELATIVE OPERATING CHARACTERISTIC CURVES

5.2.11 Probability of Indication (POI).

In some situations, it is more meaningful to compute the POI than POTD. POI is the probability that the UT signal exceeds the threshold, regardless of whether the signal is a reflection from a flaw or a result of noise. Thus, POI can be thought of as a function of POTD and PFA. The following is an ad hoc equation for computing POI:

$$POI(a) = PFA + (1 - PFA)POTD(a) \quad (11)$$

This satisfies the requirement that $POI \rightarrow PFA$ in the limit $x \rightarrow x_{no\ flaw}$ and $POI \rightarrow POD$ when signal-to-noise ratio is good ($PFA \ll 1$). Equation 11 is a rigorous result when the processes determines that noise and signal exceed the threshold and are independent. Otherwise, it is a convenient approximation that smoothly connects the limiting cases of noise for both small and large signal detection. POI is a convenient way of presenting both the PFA and POTD information in a format which is of the greatest value.

5.2.12 Summary of Terms.

In this section, a number of previously defined terms are repeated for easy reference. The following probabilities have been defined.

- POTD—The probability that threshold is exceeded by a signal from a flaw.

- POI—The probability that threshold is exceeded when a flaw is present, but not necessarily due to the signal from that flaw.
- PFA—The probability that the threshold is exceeded when no flaw is present.

In addition, other key terms are presented that will be used throughout the report. They are to be understood in the context of this work.

- Statistical Model— A description of a relationship between input parameters and a response built from empirical data.
- Physics-Based Model— A description of a relationship between input parameters and a response based on an understanding of the physics and an evaluation of mathematical expressions based on that understanding.
- Simulation—The prediction of the result of an experiment based on a physics-based model and the specification of numerical values of the input parameters.
- Test—Performing a controlled experiment in which physics-based model simulations are compared to experiment, under controlled conditions, for the purpose of determining the accuracy of the model.
- Model Verification—Establishing the accuracy of a physics-based model through a series of tests in which simulations are compared to measurements with well controlled parameters.
- Productionization of POD Predictions—The process of taking into account those effects that contribute to flaw response variability that cannot easily be quantified by simulations of well understood physical phenomena. Included are the effects of input parameters variations that are not fully controlled in the production environment.
- POD Methodology Validation—The process of establishing that the POD methodology makes predictions that properly reflect reality and provides estimates appropriate for use by the life management community.

6. SIGNAL MODELING AND VERIFICATION.

An important aspect of the new methodology is using physical models to predict the flaw signals under the influence of various inspection and material parameters, as noted by \check{Y} in the previous section. This approach reduces the empirical experimental effort and provides a basis for considering cases not covered by the experiment. In principle, any physical model for the flaw response would be satisfactory if it describes with sufficient accuracy the dependence of the flaw response on the inspection and material parameters of interest and can be accurately evaluated on computer platforms that are available to NDE personnel at the OEMs.

The following model requirements have been established to quantify these objectives:

- a. The model must be able to predict the absolute level of the flaw response, as influenced by the geometry of the part; the frequency, diameter, and degree of focusing and efficiency of the transducer; and the position of the beam with respect to the flaw. This is necessary so that the model can capture the dependence of the flaw response on \underline{x}_{NDE} and \underline{x}_{PART} .
- b. The model must be able to predict within 3 dB of careful, laboratory experiment observations. This target was established by the team based on their feeling that this was at least as good and probably better than the reproducibility of typical industrial inspections (arguments that the variability of industrial practice is as high as ± 5 dB or ± 6 dB were presented). Any greater accuracy is not supported by the industrial practice that is being described.
- c. In order to be useful to engineers in typical NDE groups of the OEMs, it is necessary that software be able to run in reasonable time on available computing platforms, which are typically personal computers, if it is to help engineers in typical NDE groups of the OEMs. Workstations are often available to individuals involved in life management but generally not to the NDE staff.

A wide variety of modeling approaches are available which satisfy the first two requirements, ranging from finite element and finite difference schemes to a various analytical approximations. The third requirement, c., must be considered in light of the more numerically intensive finite element and difference techniques and favors using analytical approximations which can decrease run times by orders of magnitude. However, the requirement of 3-dB accuracy then demands careful experimental validation of the flaw response models.

Appendix A describes the models that are used to predict the response of flat-bottom holes and synthetic hard-alpha inclusions. Also included are the results of extensive experimental verification efforts. These models do not take into account the variability of flaw response, as influenced by material microstructure. As noted in section 5.2, this is taken into account empirically by comparing the model predictions to experimental observations on sets of nominally identical reflectors. Section 9.4 presents advances in modeling these important aspects of the physics. Using these tools in future programs will further reduce the need for expensive and time-consuming experiments to determine POD.

7. RESULTS OBTAINED WITH THE NEW METHODOLOGY.

Using the new methodology described in section 5.2 and incorporating the signal modeling procedures mentioned in section 6 and described in appendix A, a number of POD (both POTD and POI) calculations for FBHs and SHAs have been performed. These, as well as some new concepts implied by ETC's approach, are described in the following subsections.

Before presenting these results, however, it may be worthwhile to review the meaning of the POTD. The procedure described in section 5.2 can be summarized as follows.

Starting with a physical model for the noise-free case, the flaw response was predicted. Then, by comparing to experimental data, which included the effects of additive noise and other microstructural influences on the flaw response (such as those that might modify the beam profile), a statistical model for the microstructurally induced deviations of the actual response from those predicted by the physical model was developed. It is the combination of these statistical and physical models which determine the PDF, from which the POTD is predicted. The statistical model was developed based on data for cases in which the flaws were clearly detected, i.e., the events were true defects. Hence, the distribution so determined is taken to be the distribution of true defects, i.e., the POTD (a).

Within this framework, the POI is estimated using equation 11. In future work, the development of a physical model, which will explicitly include the effects of noise, should be considered. This will allow the POI to be directly predicted. Section 9.4 will describe some preliminary results which lay the foundation for such an approach.

7.1 FLAT-BOTTOM HOLES.

This section briefly describes one of the predictive examples computed from the FBH experiment. The plots in the top row of figure 16 show POTD (a) and ROC curves for normal incidence inspection with a 5-MHz transducer (no. 2 in table A-2). The parameters on the curves are the threshold levels. In typical life management procedures, the threshold is selected to ensure that the required POD is realized for flaws of the critical size. Other scenarios might be encountered in other applications. Here it has been assumed that the probe is focused directly on the flaw, an assumption that will be relaxed in the discussion of the SHA inclusions. These are plots of the basic POTD, as defined in section 5.2.7. The flat-bottom holes have shown to be strong UT reflectors because they produce very high signal-to-noise ratio, even for the small 1/32" diameter (no. 1) holes (see figure 10). Therefore, in order to get ROC curves with reasonable shapes, (i.e., $POD = 1$ for no. 1 holes, except when the threshold was well above the noise), the POTD for smaller holes was predicted by extrapolating the UNDE model predictions.

The bottom row shows similar curves predicted for an inspection with the same transducer tilted 5° from normal. The POTD curves for a specific threshold are shifted to larger flaw sizes since the signals were reduced by the tilt of the probe. Quite surprising, however, was the fact that the ROC curve actually moved closer to the ideal. For a given POTD and reflection size, the false alarm rate was reduced. This indicates better inspection reliability for the tilted transducer. This result was initially felt to be counter intuitive. However, examination of the details of the calculation revealed that the noise level went down more rapidly than the signal level when the transducer was tilted, effectively improving the signal-to-noise ratio. This is not a general result, but depends on such factors as the frequency and FBH diameter. Further calculations have shown that, the larger the FBH diameter, the less pronounced is this effect since the reflectivity from the hole drops off more rapidly with angle as the hole diameter is increased. Also, this result is not achieved when the probe is tilted in the orthogonal plane.

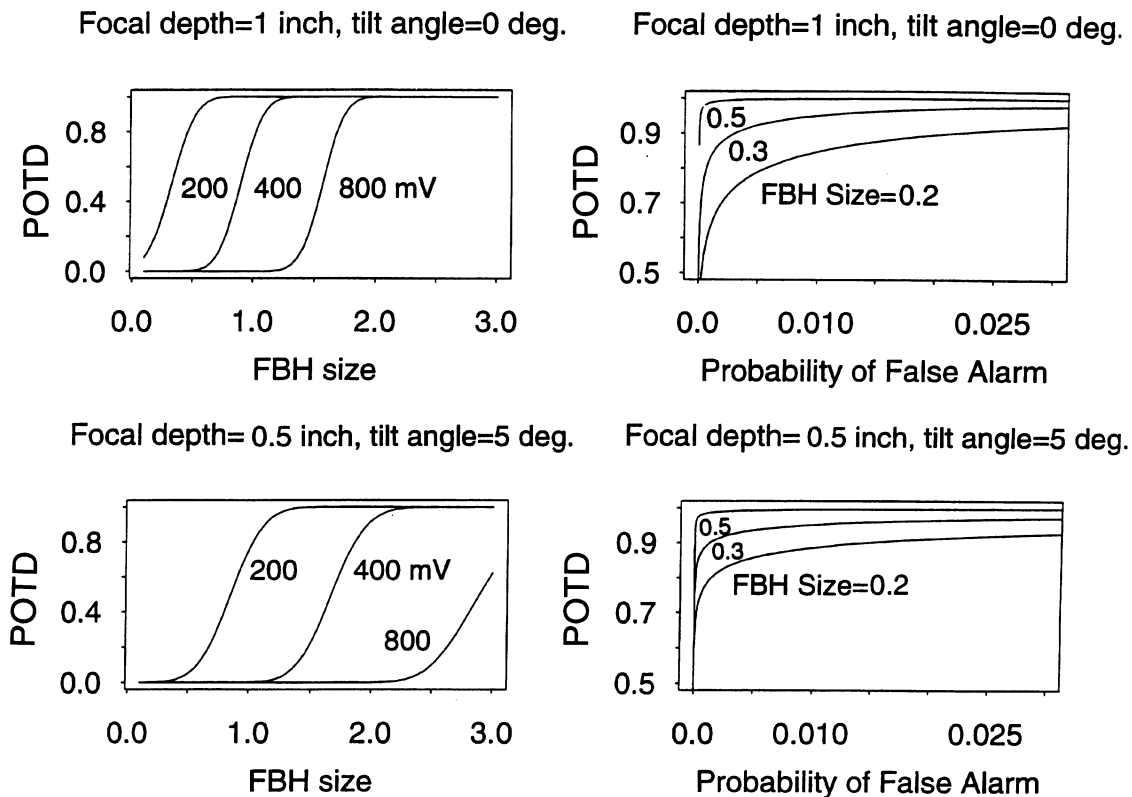


FIGURE 16. BASIC POD (a) AND ROC FUNCTIONS FOR TWO DIFFERENT SETS OF INSPECTION CONDITIONS USING TRANSDUCER NO. 2

(Flaw diameter is given in units of flat-bottom hole size in which a no. 1 flat-bottom hole is 1/64 inch. The parameters on the POD (a) curves indicate the threshold level. Flaw sizes are in units of 1/64 inch.)

It is of interest to validate these predicted improvements experimentally with a probe tilt. However, this is best done for SHAs for which the signal-to-noise ratio is closer to unity. Results will be presented for that case in the next subsection.

7.2 SYNTHETIC HARD-ALPHA INCLUSIONS.

This section briefly describes several predictive examples showing how the methodology can be used to predict POD under different inspection conditions for the SHAs (different transducers, scan plans, etc.). Experimental results also presented here will validate the methodology's predictions of improvement in POD with a probe tilt.

Figure 11 shows a C-scan as obtained with a 10-MHz transducer (no. 4 in table A-2), illuminating the sample at normal incidence and focused at the 1" depth of the SHAs. Motivated by the predicted improvement in ROC for FBHs when the probe was tilted, an additional set of C-scans were obtained. Figure 17 shows the experimental configuration. The same probe was

tilted 5° in the water. Since the SHA sample had been fabricated from a ring forging in which the microstructure, and hence the noise, is highly directional, the probe was tilted 5° with respect to the normal in each of four azimuthal directions. The resulting C-scans are shown in figure 18. Consistent with the predictions for the FBH case, the POD is improved when the probe is tilted in one plane (scans (c) and (d)) but degraded when the probe is tilted in the other plane (scans (b) and (e)).

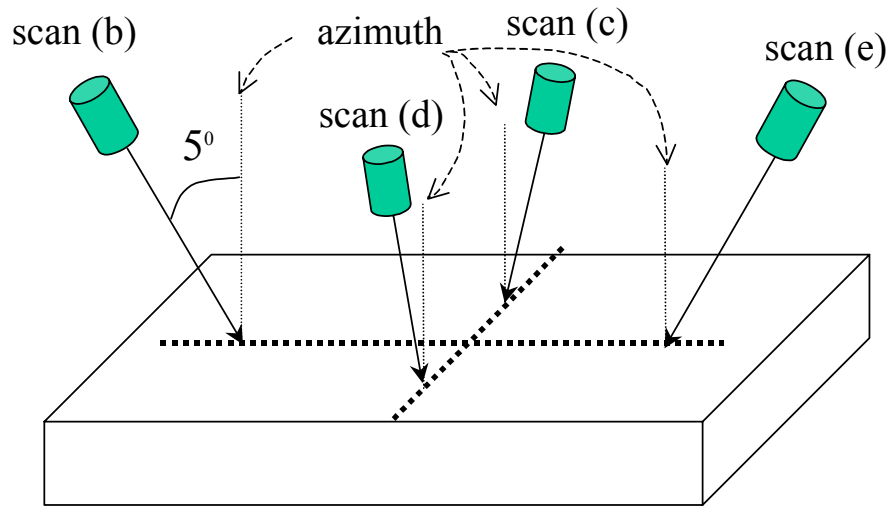


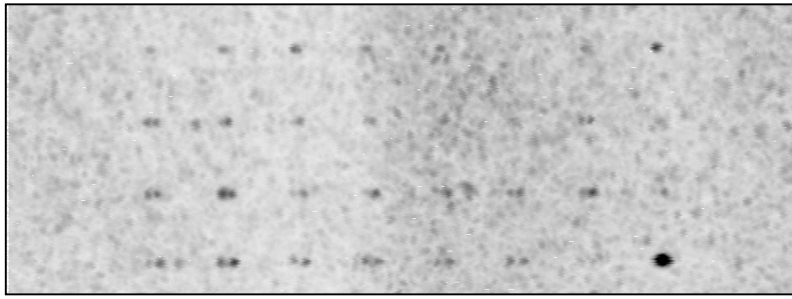
FIGURE 17. EXPERIMENTAL CONFIGURATION FOR STUDYING EFFECT OF PROBE TILT ON POTD OF SHAs

(The four azimuthal orientations of the tilted probe are indicated as scans (b)-(e). Scan (a), corresponding to normal incidence, is not shown for simplicity.)

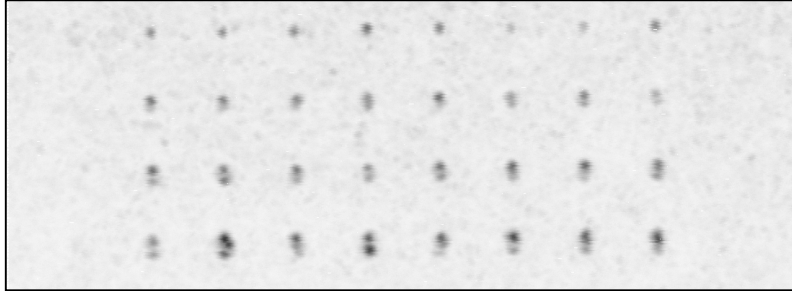
The Basic POTD predictions that used the SHA experiment data (no. 4 in table A-3) are shown in figure 19. The data was generated by probes using normal incidence and focused directly on the top surface of a nominally similar synthetic hard-alpha flaw. It is assumed that because the gate width is very small, only a thin slab of material is inspected. In contrast, figures 20 and 21 show POTD functions for 0.030" (30-mils) and 0.060" (60-mils) scan increments. In these figures, the focal depth was held constant and the beam was focused on the flaw. The figures quantify that POTD is degraded as the scan increments increase.

Corresponding to figures 20 and 21, figures 22 and 23 show ROC curves for the SHA inclusions for 30- and 60-mil scan increments. These figures allow a comparison to be made without having to adjust the threshold differences. Comparing figures 22 and 23 clearly shows how increasing the scan increment to 60 mils seriously degrades inspection capability.

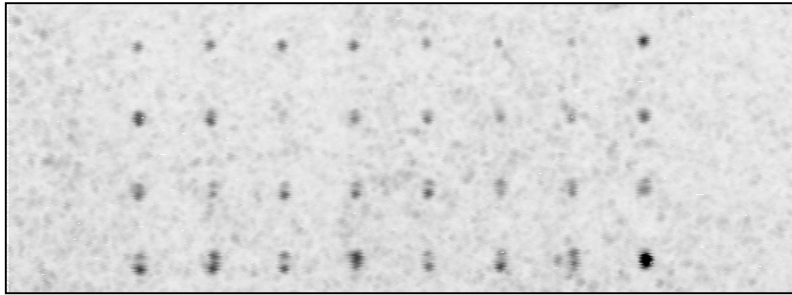
Similar to figure 20, figure 24 illustrates the difference caused by introducing a gate width that accepts signals from a slab of material 0.5" thick.



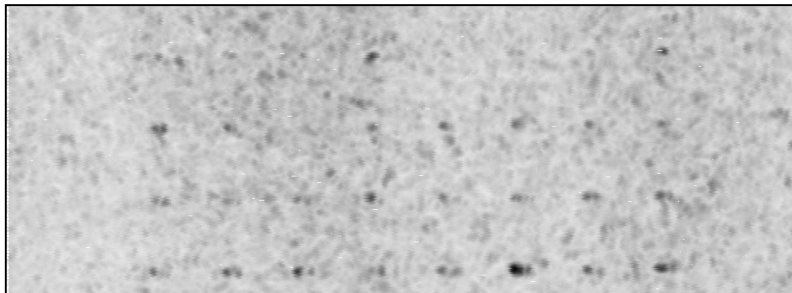
5.9% N SHA C-scan (b): 10-MHz transducer no. 4, 5 deg. tilt in water, focused at 1" depth on SHAs



5.9% N SHA C-scan (c): 10-MHz transducer no. 4, 5 deg. tilt in water, focused at 1" depth on SHAs



5.9% N SHA C-scan (d): 10-MHz transducer no. 4, 5 deg. tilt in water, focused at 1" depth on SHAs



5.9% N SHA C-scan (e): 10-MHz transducer no. 4, 5 deg. tilt in water, focused at 1" depth on SHAs

FIGURE 18. C-SCANS OF SHA BLOCK WITH 10-MHz PROBE FOCUSED IN THE PLANE OF THE FLAWS AND ILLUMINATING THE BLOCK AT A 5° TILT AND FOUR AZIMUTHAL ORIENTATIONS (see figure 17).

(The C-scan obtained at normal incidence, scan (a), was shown previously in figure 11.)

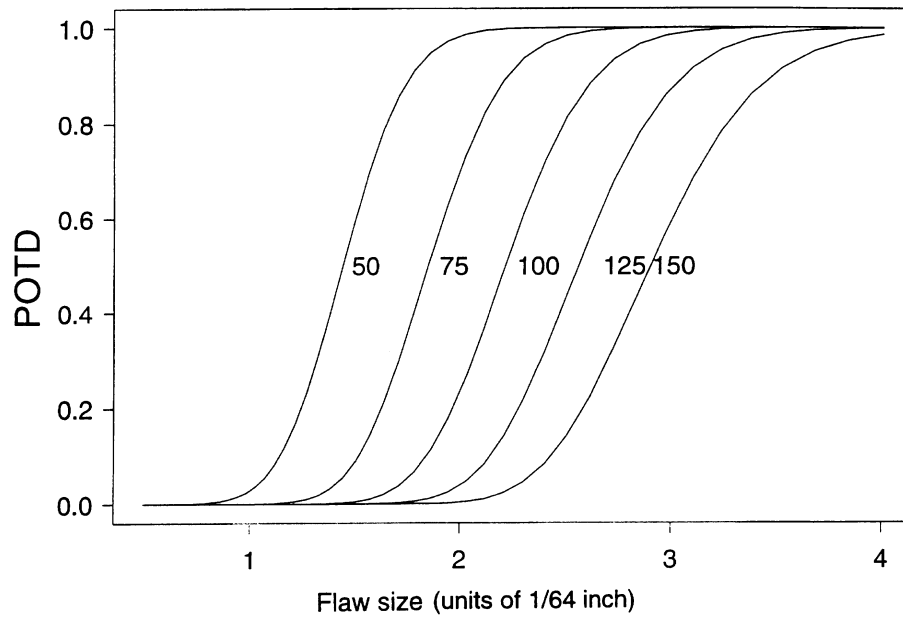


FIGURE 19. BASIC POTD FOR A 10-MHz TRANSDUCER AT NORMAL INCIDENCE, POSITIONED DIRECTLY OVER, AND FOCUSED AT THE SAME DEPTH AS THE SYNTHETIC HARD-ALPHA FLAW

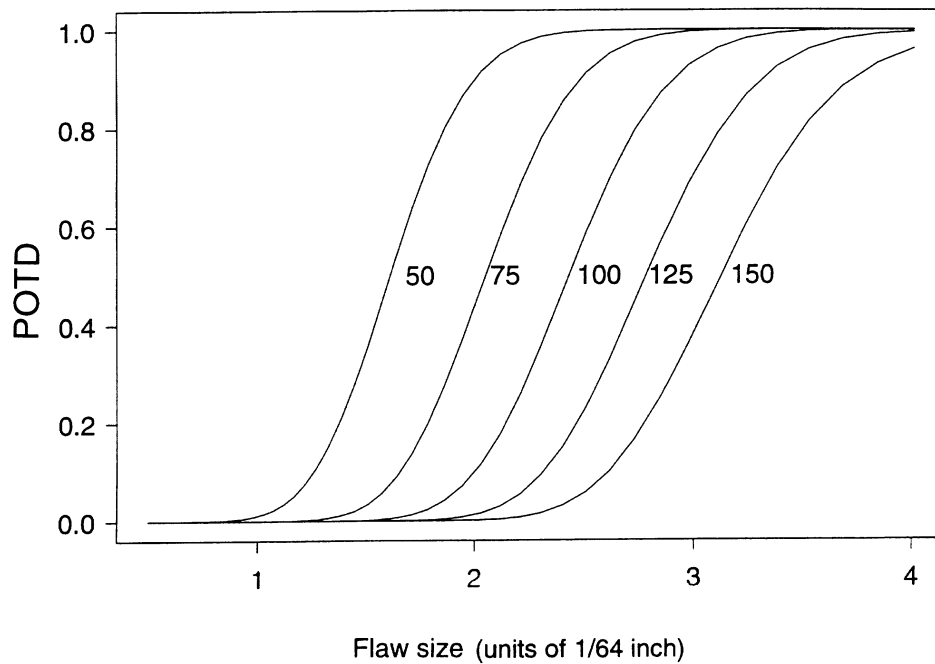


FIGURE 20. POTD FOR A 10-MHz TRANSDUCER AT NORMAL INCIDENCE, FOCUSED AT THE SAME DEPTH AS THE SYNTHETIC HARD-ALPHA FLAW, ASSUMING 30-mil SCAN INCREMENTS AND A VERY NARROW GATE WIDTH

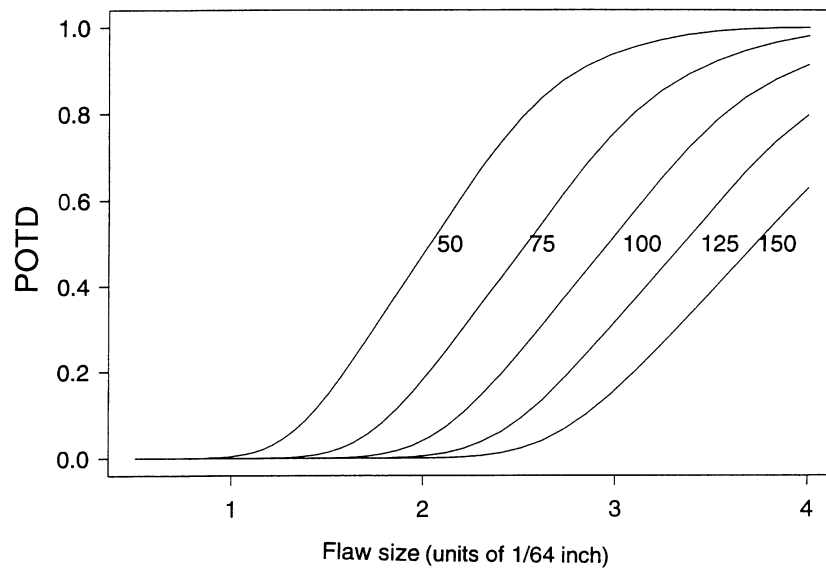


FIGURE 21. POTD FOR A 10-MHz TRANSDUCER AT NORMAL INCIDENCE, FOCUSED AT THE SAME DEPTH AS THE SYNTHETIC HARD-ALPHA FLAW, ASSUMING 0.060" (60-mil) SCAN INCREMENTS AND 0 μ s GATE WIDTH

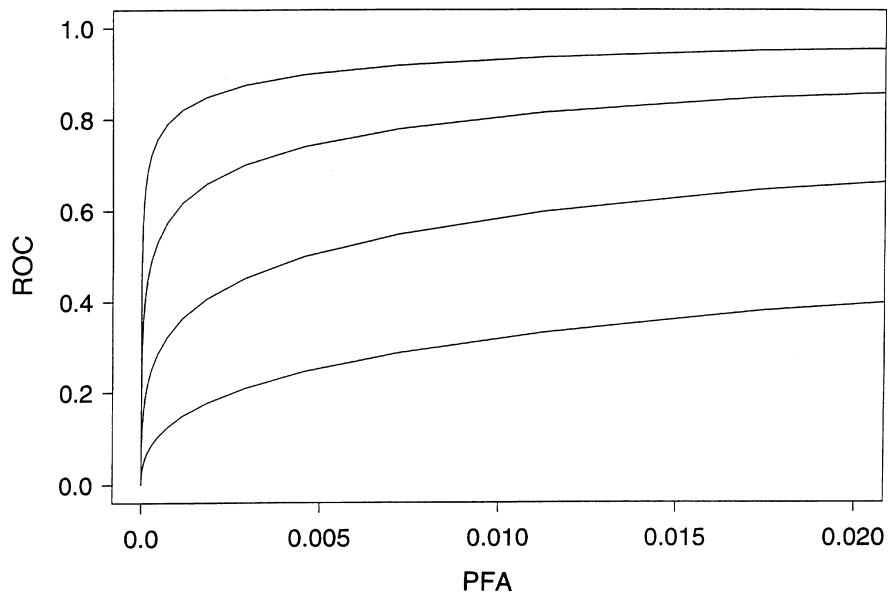


FIGURE 22. ROC FOR A 10-MHz TRANSDUCER AT NORMAL INCIDENCE, FOCUSED AT THE SAME DEPTH AS THE SYNTHETIC HARD-ALPHA FLAW, ASSUMING 0.030" (30-mil) SCAN INCREMENTS

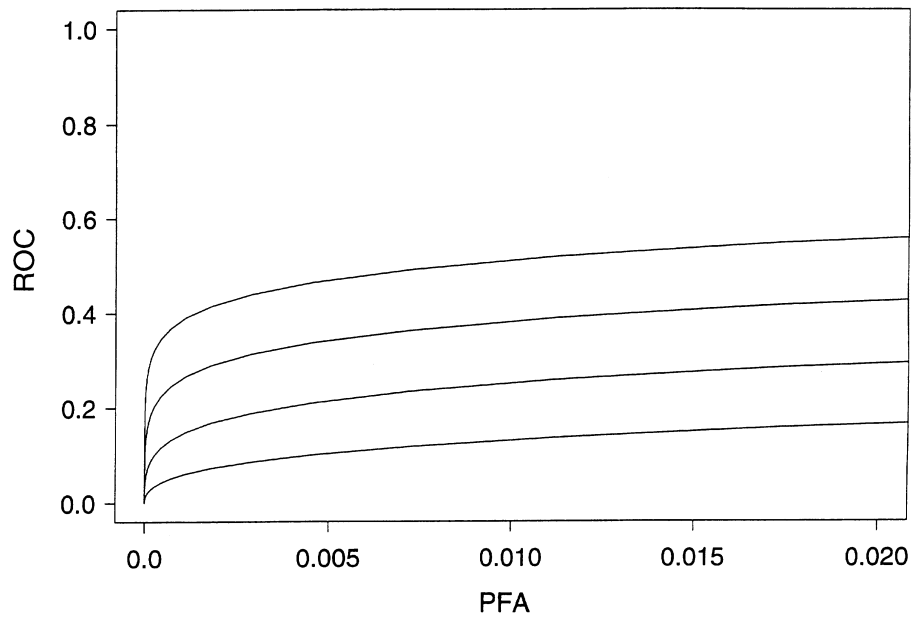


FIGURE 23. ROC FOR A 10-MHz TRANSDUCER AT NORMAL INCIDENCE, FOCUSED AT THE SAME DEPTH AS THE SYNTHETIC HARD-ALPHA FLAW, ASSUMING 0.060" (60-mil) SCAN INCREMENTS

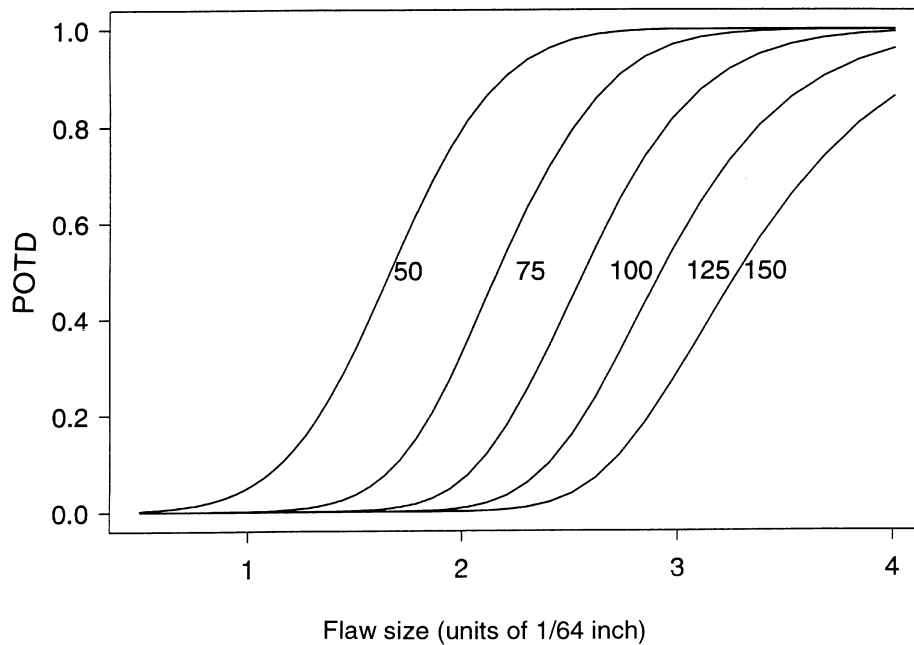


FIGURE 24. POTD FOR A 10-MHz TRANSDUCER AT NORMAL INCIDENCE, ASSUMING 0.030" (30-mil) SCAN INCREMENTS AND A GATE WIDTH OF 0.5 inch

7.3 COMPARISON OF TWO TRANSDUCERS.

Figure 25 presents a comparison of POI for a 5- and a 10-MHz transducer. The POIs were computed using normal incidence, assuming 0.020" (20-mil) scan increments and a gate width of 0.25" with the threshold adjusted so that PFA=0.02 for both transducers. The comparison clearly shows that, when inspecting with the 10-MHz transducer, there is a much higher probability of an indication occurring when flaws are present with sizes in the no. 1 to 2 range.

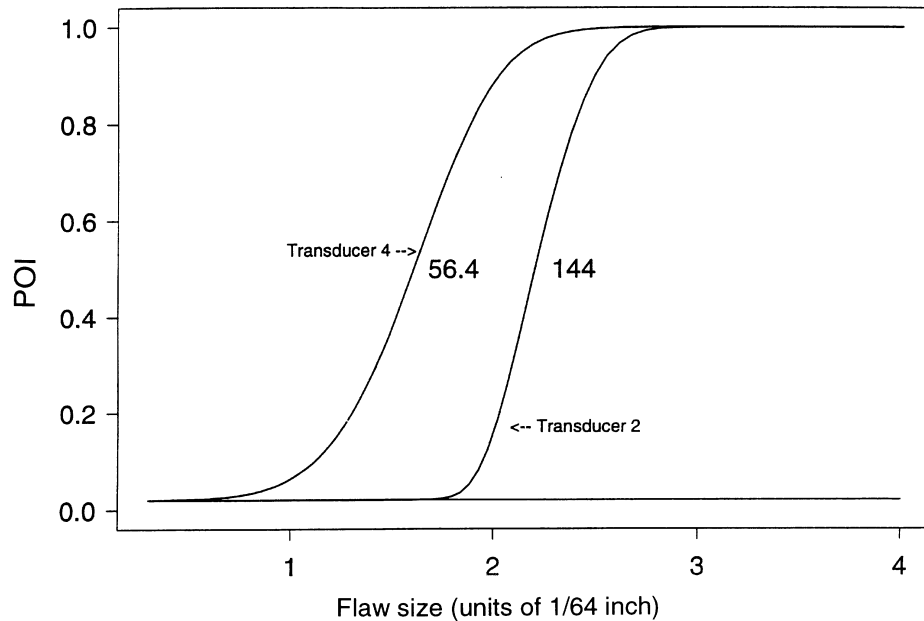


FIGURE 25. COMPARISON OF POI FOR TRANSDUCER NO. 2 (5 MHz) AND TRANSDUCER NO. 4 (10 MHz) USING NORMAL INCIDENCE, ASSUMING 0.020" (20-mil) SCAN INCREMENTS AND A GATE WIDTH OF 0.25 inch WITH THE THRESHOLD ADJUSTED SUCH THAT PFA=0.02

If one was to use a POI curve (rather than a PODT curve), one would only see whether there would be an indication of a flaw. They would not be able to say for certain if a flaw detection had occurred with the 10-MHz transducer. However, the underlying information provided by the new methodology, specifically the POTD curve not shown here, does support the conclusion of better detectability with the 10-MHz transducer. The ability to distinguish between an indication and detect of a flaw is one of the strengths of the model-based approach.

7.4 UNCERTAINTY BOUNDS.

Given the large amount of data used to estimate the distribution of generalized deviations, the dominant source of uncertainty in the predictions came from errors in the UNDE model. Since improvements were made to the UNDE model, these model predictions are estimated to be accurate within ± 3 dB, as supported by the verifications reported in appendix A. As noted in section 6, this is at least as good, if not better, than the reproducibility of typical industrial

practice. Figure 26 is an example of a calculation of uncertainty bounds, assuming this ± 3 -dB range in predicted signal amplitudes (consistent with the results reported in appendix A) and the POTD evaluation algorithm is evaluated for all possible signal predictions within a ± 3 -dB tolerance band. These bands can be viewed as the result of a graphical sensitivity analysis which outlines the potential error in the POTD predictions due to a possible inadequacy of the model. If one could confidently say that the predictions of the signals are within ± 3 dB, then they could reasonably conclude that the actual PODT is within these uncertainty bounds.

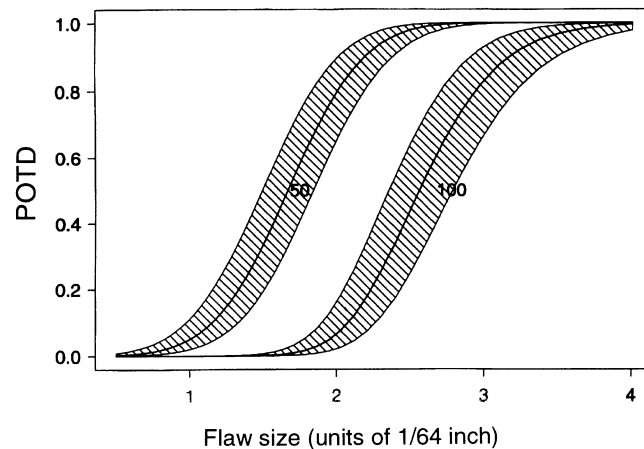


FIGURE 26. POTD FOR A 10-MHz TRANSDUCER AT NORMAL INCIDENCE, ASSUMING 0.030" (30-mil) SCAN INCREMENTS AND A GATE WIDTH OF 0.5 inch, SHOWING UNCERTAINTY BANDS FOR ± 3 dB UNCERTAINTY IN THE MODEL PREDICTIONS

8. COMPARISON OF NEW METHODOLOGY PREDICTIONS FOR FBH POD TO PREDICTIONS OF OTHER METHODS OF ANALYSIS.

The flat-bottom hole data obtained in the development of the methodology are very well-behaved in the sense that the scatter is not too large and, hence, they can be easily fit to a curve, such as a regression line. This facilitates the use of alternative methods of analysis that are usually found to be difficult to apply to detectability data from naturally occurring subsurface flaws, which exhibit considerably more scatter.

Comparison of the predictions of the new and old methodologies, based on the same data, constitutes an important verification procedure. As discussed in section 5.1, these alternative probability of detection methods may be conveniently considered as belonging to one of two groups:

- Qualitative Response: (measurements of the proportion detected for flaws of a single size or as a function of flaw size)
 - ASNT Recommended Practice
 - USAF/University of Dayton Research Institute (UDRI) PF program
 - Empirical application of the Relative (or Receiver) Operating Characteristic

- Quantitative Response: (measurements of signal amplitude as a function of flaw size)
 - USAF/UDRI \hat{a} versus a (a -hat versus a) program
 - GEAE Effective Reflectivity method (original and modified versions)

This section provides examples of applications of several of these methods to the FBH data. In all of these applications, it has been assumed that the size of the FBHs is exactly determined by the nominal values, e.g., all no. 2 FBHs are treated as having a diameter of 2/64", etc. (It must be remembered that this prior knowledge of the flaw sizes is not available when studying naturally occurring subsurface flaws.)

Attempts were also made to conduct similar studies based on the SHA data. However, sensible results were not obtained with the existent methodologies. This illustrates the greater versatility of the new methodology.

8.1 QUALITATIVE RESPONSE (HIT-MISS) METHODS.

8.1.1 American Society of Nondestructive Testing Recommended Practice.

Although the full title of reference 6 reflects the aerospace industry background to this document, its statistical approach is not limited to specific aerospace applications. The goal was to determine the limiting flaw size that could be detected given the probability of detection and the percentage of confidence in that probability. The hit-miss approach focuses attention on flaws of a single size or size interval.⁹ This is based on measuring the response of the proportional group of flaws that is detected by an inspection.

The formulas of this method use the percentiles of the F-distribution [23] as the basis for calculating the one-sided confidence limits. In its simplest form, the ANST Recommended Practice identifies pairs of values for trials (i.e., sample sizes) and successes (detections) necessary to establish a 90% POD at eight different confidence levels, ranging from 50% to 99.9%. For example, detection of seven flaws out of a sample of seven is enough to demonstrate 90% POD at 50% confidence, whereas it is necessary to detect 29 out of 29 to demonstrate 90% POD at 95% confidence. If a few flaws are missed, the total number of trials increases rapidly, i.e., demonstrating 90% POD at 95% confidence requires the detection of 45 out of 46 flaws or 59 out of 61.

The F-distribution may give a more detailed analysis that will help to understand the properties of specific data sets. For example, it is possible to evaluate, as a function of the number of trials and successes, the POD for a specific confidence level or the confidence level for a specific POD. The former technique has been applied to the three sets of no. 1 FBH data; results expressed in terms of the probability of detecting no. 1 FBHs are shown in figure 27. There are six curves, one for each of three transducers for a 50% confidence level and a 95% confidence level. One would expect (1) that the POD for a given size reflector would decrease as threshold increases, (2) that the POD would be lower if a higher confidence is required, and (3) that the

⁹ If flaws of different sizes are used, the procedure leads to certification "at the...flaw size interval," and the practice illustrates POD plotted at the center of such an interval (see figure 4-1 of reference 6). This appears to be dangerously nonconservative since it tends to over represent the POD at the lower end of such an interval.

POD would be higher for a transducer producing a stronger signal. All of these general expectations are born out by these specific calculations. As the inspection (accept/reject) threshold is increased from 300 to 500 mV, the POD for transducers 1 and 2 falls in steps from 0.957 at 50% confidence and 0.829 at 95% confidence to zero. The POD for transducer 3, because of its higher response values, stays at the initial values (and then falls to zero as the threshold is increased from 620 to 890, which is off of the scale of this figure).

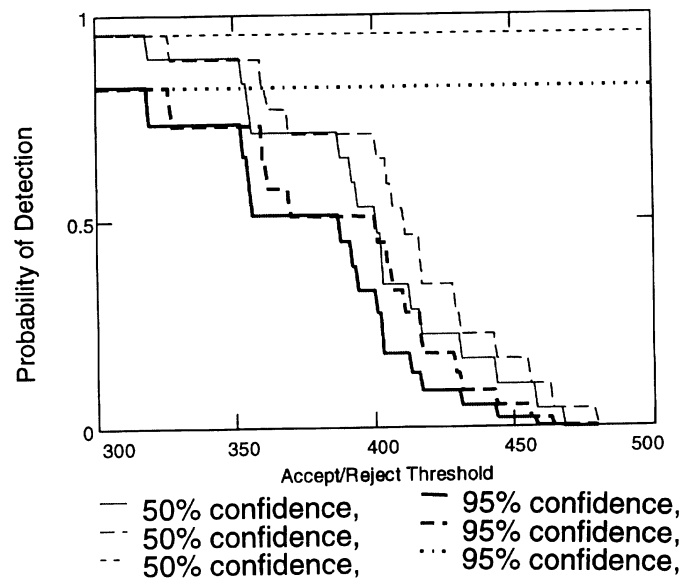


FIGURE 27. AMERICAN SOCIETY FOR NONDESTRUCTIVE TESTING
RECOMMENDED PRACTICE POD FOR NO. 1 FBHs AT TWO CONFIDENCE LEVELS AS
A FUNCTION OF THRESHOLD LEVEL

(The solid, dashed, and dotted curves refer to transducers 1, 2, and 3 respectively.)

This procedure is not well-suited to plotting POD as a function of flaw size for this data set¹⁰. For the purposes of comparison with other methods of analysis, the values of POD for a threshold of 400 mV are listed in table 2.

TABLE 2. PROBABILITY OF DETECTING NO. 1 FBHs WITH A 400-mV THRESHOLD

	Transducer No. 1		Transducer No. 2		Transducer No. 3	
Confidence	0.50	0.95	0.50	0.95	0.50	0.95
POD	0.531	0.333	0.714	0.516	0.957	0.829

¹⁰ By grouping the data into intervals that include more than one size of FBH, it is possible to show that POD increases with increasing FBH size – for example, with a threshold no higher than 318 mV, 32 out of 32 FBHs in an interval, including no. 1 and 2 sizes are detected; for PODs of 0.979 and 0.911 at 50% and 95% confidence, respectively, values that are higher than the 0.957 and 0.829 derived from detection of 16 out of 16 no. 2 FBHs alone. However, this increase vanishes as soon as the threshold is raised high enough for a single no. 1 FBH to be missed. It is then better to confine attention to separate intervals containing either no. 1 or 2 FBHs.

8.1.2 USAF/UDRI PF PROGRAM.

This software program, written for the USAF at the UDRI [4, 7, 24, and 25], is intended to serve as an analysis of pass/fail data, i.e., for detection data, distinguishing only by whether known flaws are detected (pass or hit) or not detected (fail or miss). The PF program assumes a lognormal¹¹ model for the $POD(a)$ function, where a is the flaw size and calculates maximum-likelihood estimates of the parameters of this model. For this program to run properly, the data set must include small always undetected flaws, large always detected flaws, and an intermediate range in which some flaws are detected and others are missed [7].

For the present FBH data, 77 threshold values were tried at 50-mV intervals from 300 to 4150 mV, using data from transducer no. 2. The PF program failed to determine parameters for the $POD(a)$ model for 68 of these thresholds. It ran, with limited success, 1150 mV, 3850 mV, and 7 thresholds in the range between 2250 and 2550 mV. This generated model parameters from which mean, 50% confidence POD could be calculated as a function of flaw size. Examples of POD curves that were generated from successful conditions are shown in figure 28. The output did not include 95% confidence estimates, “due to inadequate fit to the POD model”¹².

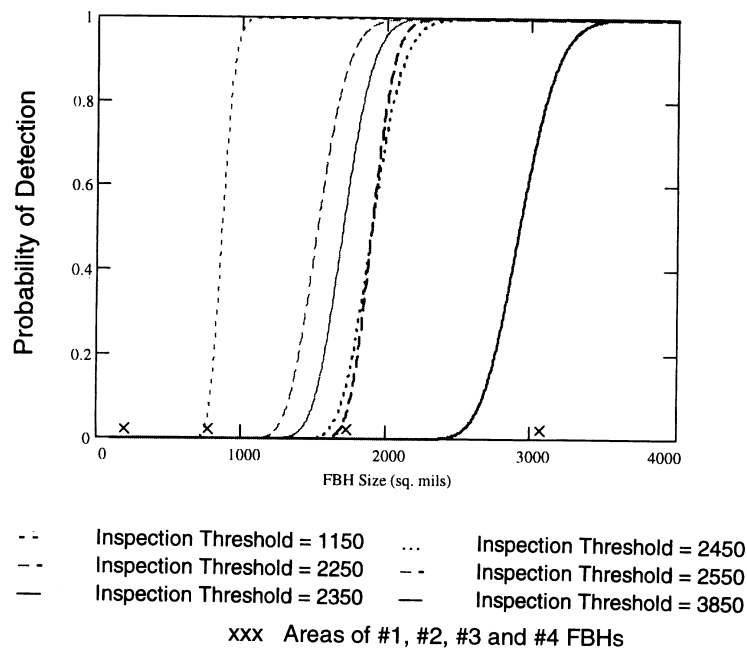


FIGURE 28. USAF/UDRI PF ANALYSIS POD FOR PLANAR VOIDS BASED ON PF ANALYSIS (TRANSDUCER NO. 2) (Note: 1000 sq. mils = 0.001 in.²)

¹¹ Earlier versions of the PF software may have used the log-logistics function, which gives almost identical results.

¹² Berens cautions that this indicates that “the log odds model is not acceptable for the data set” and that under these circumstances “the parameters of the POD function are output for reference.”

Table 3 compares results from the analyzed data of transducer no. 2 with the ASNT Recommended Practice and the PF program. Because of the restriction imposed by the ASNT Recommended Practice, this comparison was limited to listing the values of the probability of detecting no. 3 FBHs (i.e., planar voids of area 0.001726 in.²). For a given threshold, (and specifically for this set of data), the PF program indicates slightly higher values of mean POD, but the ASNT Recommended Practice yields results over a wider range of threshold and confidence conditions.

TABLE 3. PROBABILITY OF DETECTING NO. 3 FBHs WITH TRANSDUCER NO. 2

Threshold (mV)	POD From PF		POD From ASNT	
	50% Confidence	95% Confidence	50% Confidence	95% Confidence
2200	N/A	N/A	0.957	0.821
2250	0.883	N/A	0.836	0.656
2300	0.804	N/A	0.775	0.583
2350	0.559	N/A	0.531	0.333
2400	0.379	N/A	0.347	0.178
2450	0.128	N/A	0.103	0.023
2500	0.128	N/A	0.103	0.023
2550	0.058	N/A	0.042	0.003
2600	N/A	N/A	0.000	0.000

8.1.3 Empirical ROC Approach.

The relative (or receiver) operating characteristic [12] was originally conceived in conjunction with the development of signal detection theory as a way to present data necessary to evaluate the process of choosing between two alternative responses. The ROC format has also found application, without particular regard to its detection theory origins, as a convenient way to simultaneously display POD and PFA data. When it is presented this way, it has sometimes been distinguished as an Empirical ROC application.

Since the current FBH data set is confined to detection data and contains no information pertinent to PFA, it is not a useful application for the ROC approach.

8.2 QUANTITATIVE RESPONSE METHODS.

8.2.1 USAF/UDRI \hat{a} Versus a Program.

This software program, written for the USAF at the UDRI [4, 7, 24, and 25], is intended to analyze detection data in the form of output signals—the amplitudes of which are (increasing) functions of the flaw size. By convention, the response is designated as \hat{a} or \hat{A} (or “A-hat”) and the size is designated as a or A . The program assumes a linear dependence of $\ln(\hat{a})$ on $\ln(a)$ and calculates maximum-likelihood estimates of the parameters β_0 , β_1 , and ϵ for the linear regression equation:

$$\ln(\hat{a}) = \beta_0 + \beta_1 * \ln(a) + \varepsilon \quad (12)$$

The program further assumes that the residuals, ε , are normally distributed with a mean of zero, and the standard deviation, $\sigma(\varepsilon)$, is independent of a (i.e., the residuals are homoscedastic). The $POD(a)$ function then has the form

$$POD(a) = \Phi[\{\ln(a) - \mu\} / \sigma] \quad (13)$$

where Φ is the cumulative normal distribution,

$$\mu = \{\ln(a_{dec}) - \beta_0\} / \beta_1, \quad (14)$$

$$\sigma = \sigma(\varepsilon) / \beta_1, \quad (15)$$

and a_{dec} is the decision threshold.

The program itself has a wide range of input conditions, which only requires the slope of the regression line to be positive. The user is left to test for linearity, normality, and homoscedasticity of the residuals. Berens has recently suggested [25] using the statistical tests for these properties that are provided by Minitab, Release 10.5 *Xtra*, at a 0.1 significance;¹³ examples of applications of these tests are included in table 4.

TABLE 4. MINITAB RELEASE 10.5 *Xtra* TESTS APPLIED TO \hat{a} REGRESSION FOR TRANSDUCER NO. 2

Property	Test	p
Linearity of regression	XLOF	0.000
Normality of residuals	Anderson-Darling	0.101
	Ryan-Joiner (Shapiro-Wilk)	> 0.1
	Kolmogorov-Smirnov	> 0.15
Homoscedasticity of residuals (versus log(Area))	Bartlett	0.000
	Levene	0.001

Since the program does not apply a model for the response, the user is free to make arbitrary selections of parameters to be used for a and \hat{a} . As figure 29 demonstrates, for this set of data, plotting mV response versus either flaw diameter or flaw area produces a near-linear graph in the log-log coordinates that the program uses¹⁴. Area was chosen as the basis for the current analysis, since the near-linearity of the graph of mV versus area in linear-linear coordinates will be important to the subsequent Effective Reflectivity analysis (whereas the graph of mV versus diameter is distinctly nonlinear).

¹³ There is, as yet, no industrywide agreement on the appropriateness of these tests or the recommended significance.

¹⁴ Dimensions used for diameter and area are mils (1 mil = 10^{-3} in.) and sq. mils. respectively (1 sq. mil = 10^{-6} in²).

Figure 30 shows that the data are not quite linear. However, the regression line provided by the program is, above, all the measured data for the no. 2 FBH. The average observer would probably accept these data as quasi-linear, but the Minitab tests (suggested by Berens) resoundingly reject the hypothesis of normality and homoscedasticity but fails to reject the hypothesis of lognormality of the residuals. Results are listed in table 4, and the residuals are plotted on a cumulative probability chart in figure 31. It can be speculated that the rejection of normality is related to the fact none of the 16 no. 2 FBH points is above the regression line.

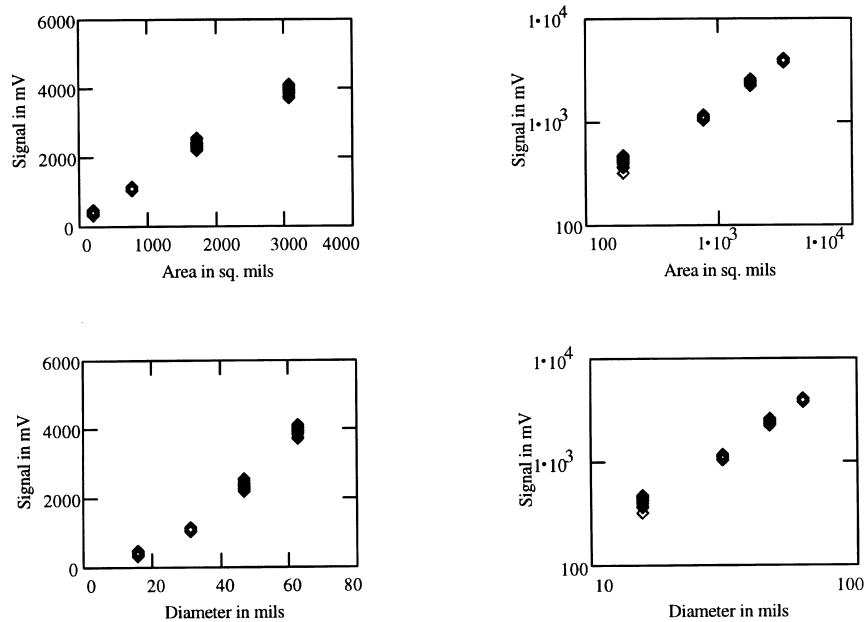


FIGURE 29. COMPARISON OF SCATTER-PLOTS OF SIGNAL RESPONSE FOR TRANSDUCER NO. 2 VERSUS FBH AREA OR FBH DIAMETER IN LINEAR OR LOGARITHMIC COORDINATES (Note: 1000 mils = 1 inch)

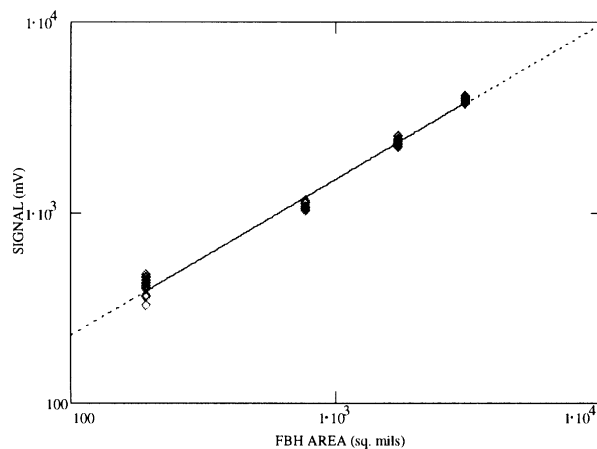


FIGURE 30. COMPARISON OF TRANSDUCER NO. 2 DATA WITH \hat{a} REGRESSION LINE

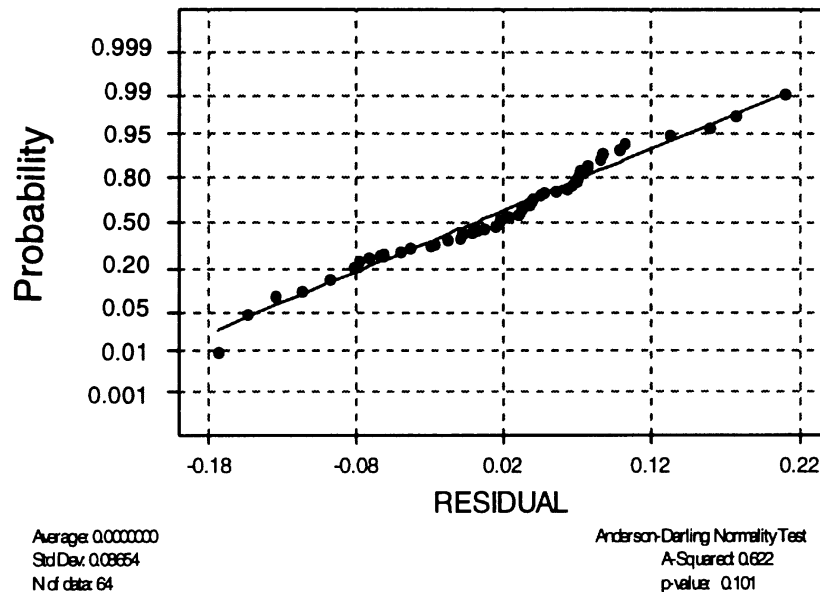


FIGURE 31. CUMULATIVE NORMAL PROBABILITY PLOT OF THE RESIDUALS FROM THE \hat{a} REGRESSION OF LOG (mV) ON LOG (AREA) FOR TRANSDUCER NO. 2

It is suggested in MIL-HDBK-1823

- a. that if deviations from the assumptions of linearity, normality, and homoscedasticity cannot be corrected it suffices to note them [26].
- b. that the range of flaw sizes considered may need to be restricted in order to ensure adequate conformance [27].
- c. that “as a minimum, these assumptions must be subjectively evaluated by visual examination of a plot of $\log \hat{a}$ vs. $\log a$ ” [28].

Consistent with the first of these suggestions, the non-conformances have been noted (estimates of POD obtained from the program should, therefore, be viewed as somewhat questionable validity). Results from the analysis for data from transducer no. 2 are shown in figure 32 for the set of six thresholds used in the signal distribution analysis, forming part of the proposed ETC approach to POD (see figure 16) and for the lowest of the viable PF thresholds shown in figure 28. For the last of these thresholds, the program indicates slightly smaller values of flaw area than does the PF program (for example, 826 sq. mils versus 950 sq. mils at 0.9 POD or 631 versus 783 at 0.1 POD). To facilitate a comparison with the signal distribution (see figure 16) analysis, results from figure 32 have been replotted in figure 33, using FBH diameter in 64ths of an inch as the abscissa.

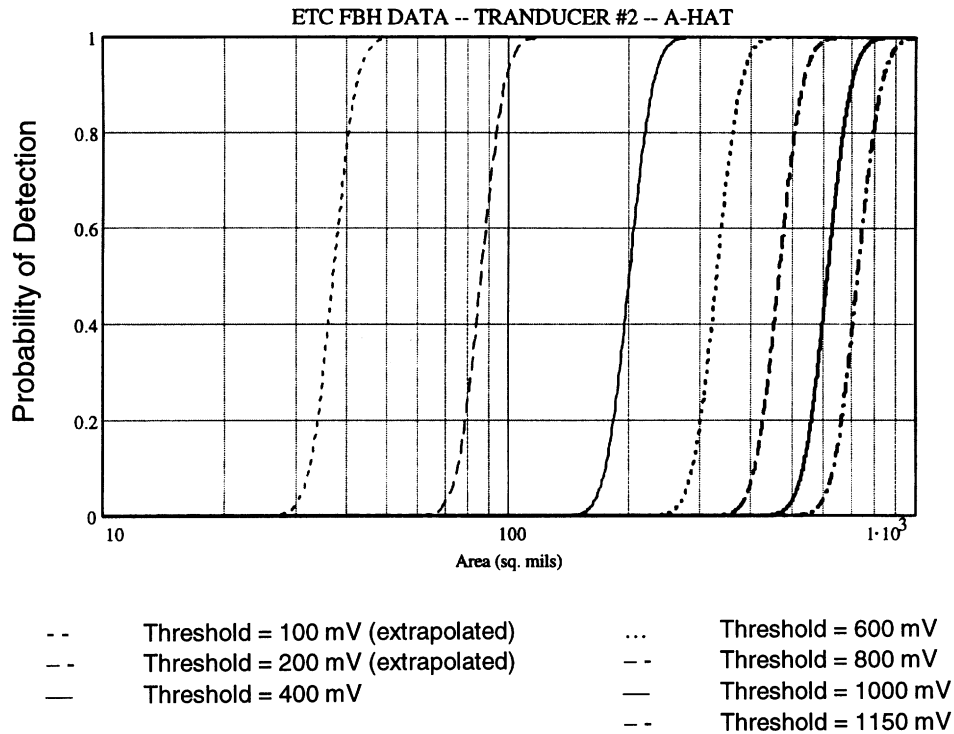


FIGURE 32. USAF/UDRI \hat{a} ANALYSIS OF POD VS FLAW AREA FOR PLANAR VOIDS

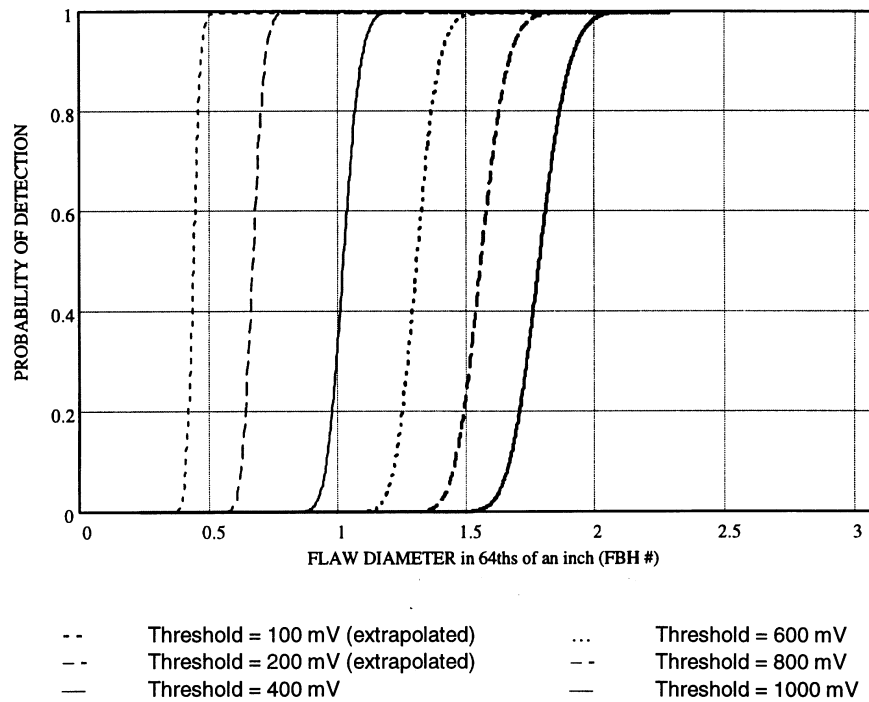


FIGURE 33. USAF/UDRI \hat{a} ANALYSIS OF POD VS FLAW DIAMETER FOR PLANAR VOIDS

Since using inspection thresholds below 380 mV requires extrapolation outside the range of experimental data, less emphasis should be placed on the POD curves for thresholds of 100 and 200 mV than for the higher thresholds displayed in figures 32 and 33. It should also be noted that, if extrapolated to even lower threshold values, the regression line of figure 30 implies a finite signal for a vanishingly small reflector. Among the possible causes of this effect is the presence of noise.

8.2.2 GEAE Effective Reflectivity Method.

This method was developed specifically to deal with POD estimations for ultrasonic inspections which consist of small numbers of data points exhibiting considerable scatter—i.e., data for which use of a simple linear regression technique (one unguided by a physical model), such as used in the \hat{a} program—would be likely to result in arbitrary and unrealistic prediction of the dependence of response on flaw size. A conceptual sketch of how a model can stabilize the regression is presented in figure 6.

The method uses a simple mathematical flaw response model to provide a means to interpret the data from natural flaws in a physically plausible manner. The commonly used linear area-amplitude characteristic [9]¹⁵ of the FBH¹⁶ provides a reasonable first approximation to the response of defects in the transducer field. Although this technique was not developed with regression in mind, it is equivalent to linear regression of the signal on the area (in linear-linear coordinates) with an intercept set at zero by the assumed model (again, see figure 6).

8.2.2.1 Effective Reflectivity Analysis: Initial Method.

As originally conceived [8], this method involves:

- a. measuring the response from a set of FBHs at various depths using the same instrumentation and scanning parameters as intended for actual product inspection.
- b. using the FBH data to calculate Equivalent FBH (EFBH)¹⁷ areas for indications found during product inspection.
- c. performing a detailed metallographic examination of the regions from which selected indications originated with a goal of establishing the precise volumetric dimensions of each defect.
- d. calculating the ratio of the EFBH defect size predicted by the model to the measured size; this ratio is called the Effective Reflectivity, R_e , and properties of the distribution of values of R_e are used to calculate POD taking advantage of the lognormality that has usually been found.

¹⁵ Strictly applicable only well into the far field of a single-frequency unfocused transducer; see reference 9.

¹⁶ Several alternatives for which mathematical models were readily available (such as cylinders and spheres – see reference 9) were evaluated, but the far-field approximation to the FBH was found to be most effective.

¹⁷ i.e., the area of a hypothetical FBH, normal to the sound beam, at the same depth as the indication that would give the detected indication amplitude.

- e. identifying the smallest FBH that could be detected at the selected inspection threshold.
- f. dividing this FBH size by various percentiles of the R_e distribution in order to obtain the estimates of the size of the real defects corresponding to various values of POD (for example, if the value corresponding to 0.1 cumulative probability is used, the resulting area is taken as a 50% confidence estimate of the size of defect that could be detected with 90% POD since 90% of the defects would have an R_e at least this large).
- g. calculating lower one-sided 95% confidence estimates; this is conditional on that the R_e values being distributed normally (or lognormally) [8].

The current set of data has no independent FBH calibration of the type previously described in step a. Instead, Minitab has been used to fit a regression line with zero intercept, in mV versus area coordinates, consistent with the assumed linear area-amplitude model. It may be seen from figure 34 that the regression for transducer no. 2 data obtained in this way is not a particularly good fit. It is, therefore, not surprising that the R_e distribution is irregular and, as shown in figure 35, is neither acceptably normal nor lognormal.

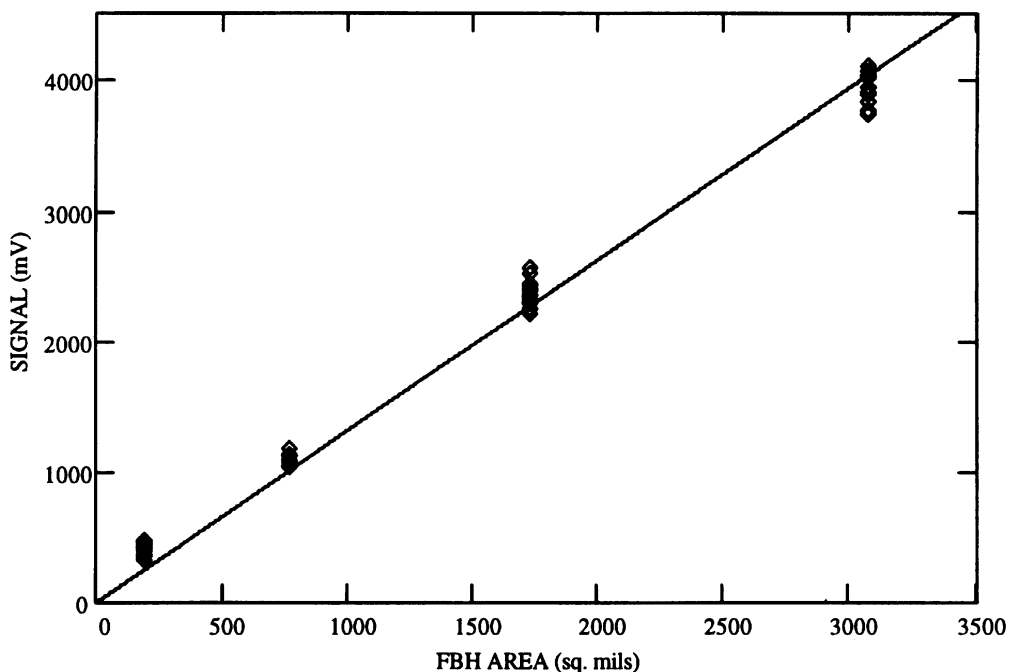


FIGURE 34. COMPARISON OF DATA WITH R_e REGRESSION LINE
(TRANSDUCER NO. 2)

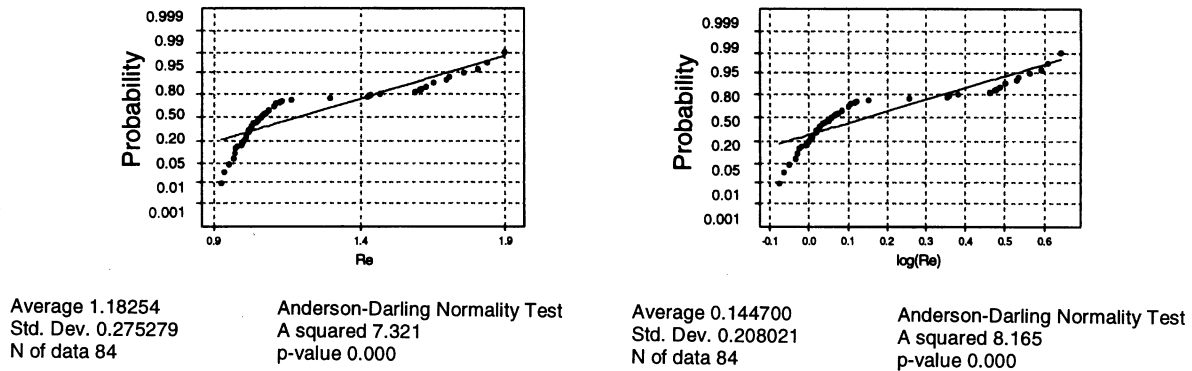


FIGURE 35. CUMULATIVE NORMAL PROBABILITY PLOTS FOR R_e AND $\log(R_e)$ (TRANSDUCER NO. 2)

Although it is advantageous to use the properties of a standard distribution (where this is justified), the Effective Reflectivity method can be used empirically, and the POD, which is derived from probability values and R_e , can be taken from figure 35. Results from the transducer no. 2 data are shown in figure 36. The EFBH corresponding to each of the threshold values is calculated from the regression relationship of figure 35. The flaw sizes are then calculated by dividing these limiting FBH areas by the appropriate values of Effective Reflectivity, as described in step f. Finally, these flaw areas were converted to planar disk diameters, expressed in 64ths of an inch, and were compared with other sets of POD data. These results are generally compared to those from the A-hat analysis (cf. figure 33) although they show more irregular dependence of POD on flaw size.

8.2.2.2 Effective Reflectivity Analysis: Initial Method, Adjusted for Noise.

Gilmore [29] has hypothesized that the area-amplitude linearity assumed by the FBH model were being concealed by the effects of the additive noise so that the observed signal amplitudes were actually signal plus noise. Analyses of this data, using the assumptions of the linear area-amplitude relationship with a constant noise added to each FBH signal, which is independent of the FBH diameter, results in the conclusion that the hypothesized noise is 173 mV¹⁸. Figure 37 shows that the fit of the regression line clearly improved by subtracting the assumed noise amplitude from each FBH response (although it is accepted as linear by the Minitab XLOF test at only 0.003 significance!). Figure 38 shows that the residuals are still neither normal nor lognormal, but are distributed more symmetrically than the residuals shown in figure 35.

¹⁸ This process is equivalent to performing linear regression in linear-linear coordinates, on the original data, and allowing the regression analysis to select a nonzero intercept.

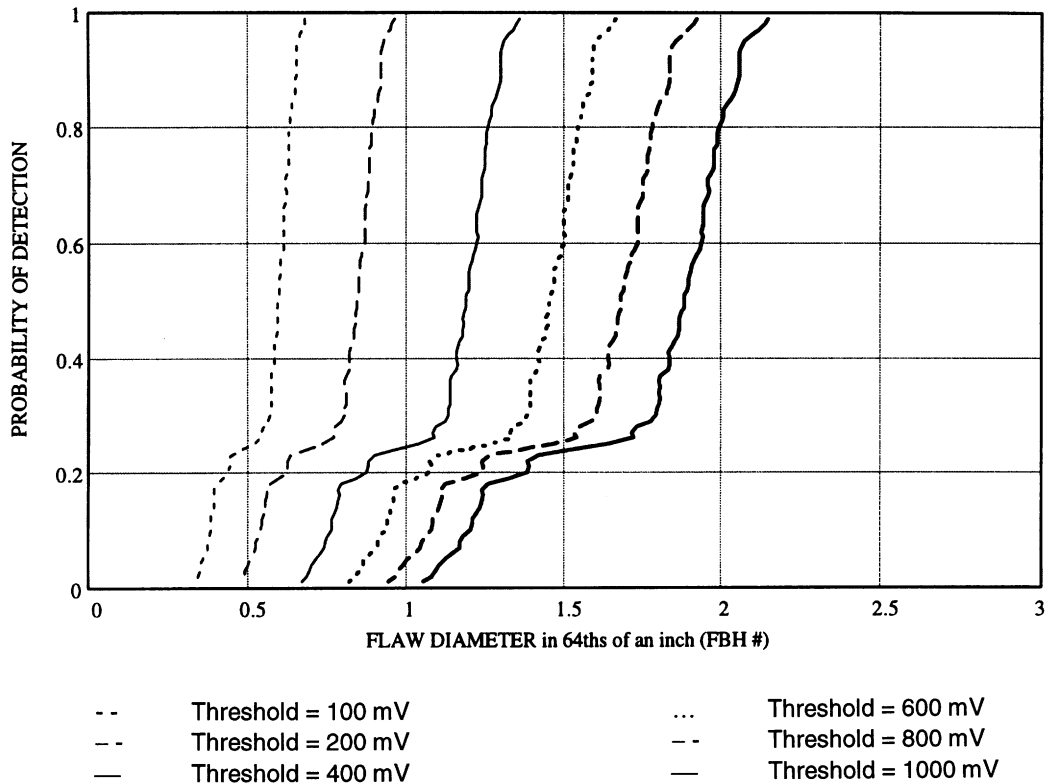


FIGURE 36. GE EFFECTIVE REFLECTIVITY ANALYSIS POD FOR PLANAR VOIDS, LINEAR AREA-AMPLITUDE MODEL (TRANSDUCER NO. 2)

It would be possible (and, perhaps, appropriate) to extract R_e values from figure 37 in the same way that was done for figure 35. To illustrate the type of POD curve given by the Effective Reflectivity approach in cases where a standard distribution is justifiable, the data in figure 37 will be treated as normally distributed, with location (0.9958) and scale (0.0938) consistent with the straight line plotted in figure 38.

POD values are generally similar to those in figures 33 and 36 and are shown in figure 39; although POD for the 200-mV threshold is predicted to be better, i.e., figure 39 shows that flaws of smaller size would be detected at this threshold that is the case for figures 33 and 36. The 100-mV threshold was not used in constructing figure 39, since it was below the hypothetical noise level of 173 mV noted two paragraphs above. In the presence of such noise, a 100-mV threshold would lead to a rejection on each inspection opportunity since the noise would be 73% higher than this threshold. The POD predicted in this case would be meaningless. However, since this approach is nothing more than a hypothesis, there is no direct experimental support.

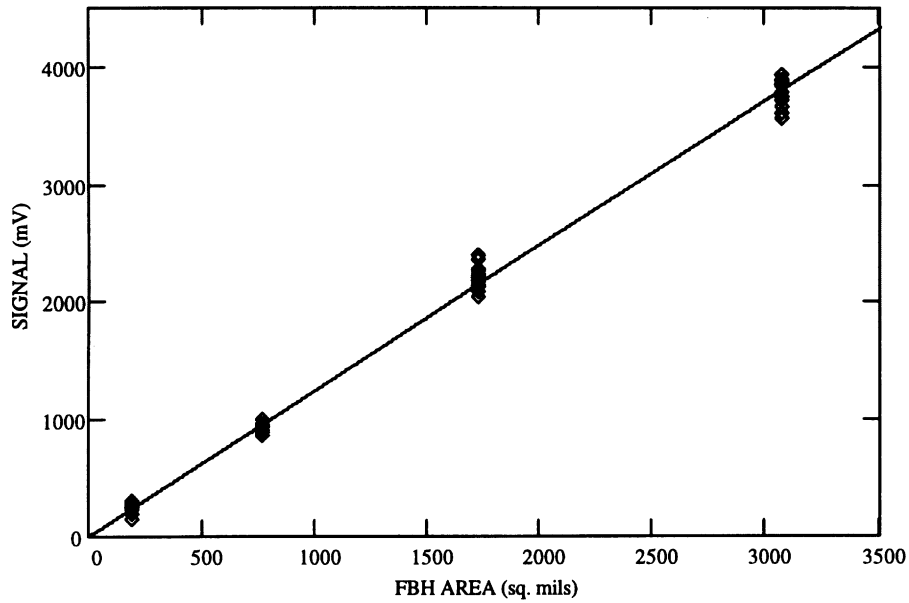
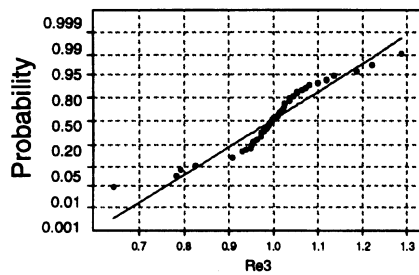
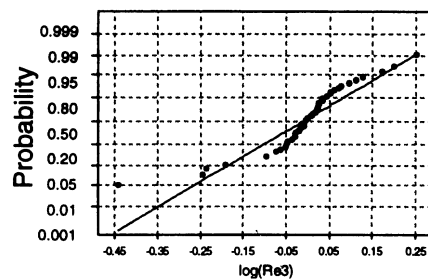


FIGURE 37. COMPARISON OF NOISE-ADJUSTED DATA WITH R_e REGRESSION LINE (TRANSDUCER NO. 2)



Average 0.996776
Std. Dev. 0.0987667
N of data 84

Anderson-Darling Normality Test
A squared 2.264
p-value 0.000



Average -0.0066617
Std. Dev. 0.0966794
N of data 84

Anderson-Darling Normality Test
A squared 2.688
p-value 0.000

FIGURE 38. CUMULATIVE NORMAL PROBABILITY PLOTS FOR NOISE-ADJUSTED R_e AND $\log(R_e)$ (TRANSDUCER NO. 2)

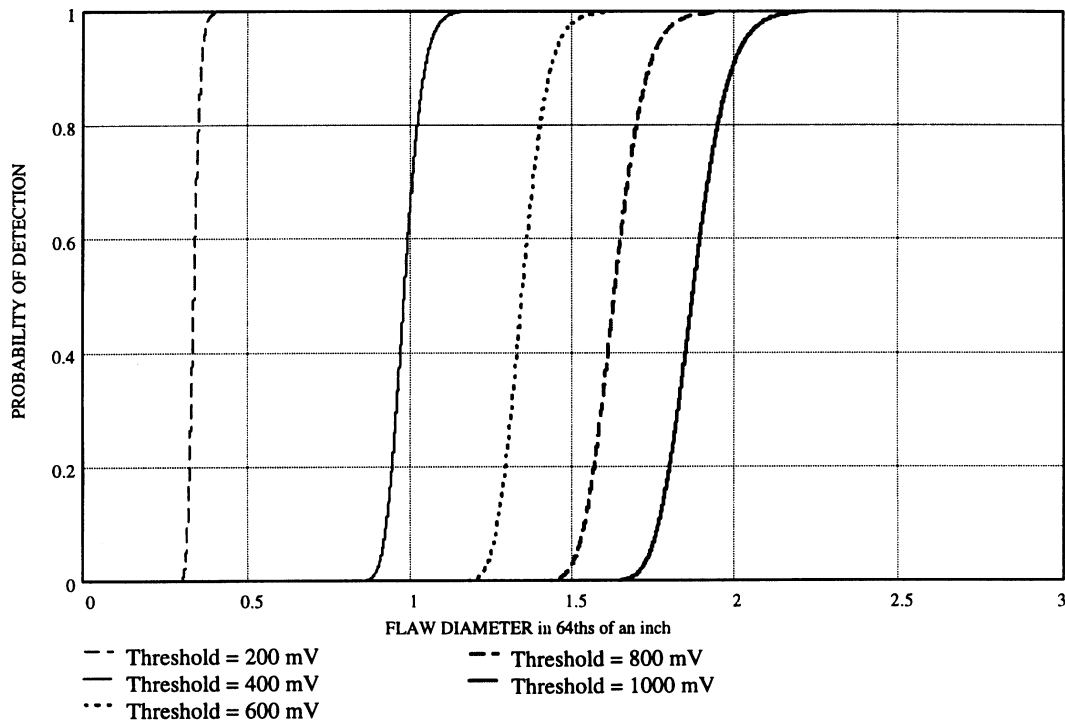


FIGURE 39. GE EFFECTIVE REFLECTIVITY ANALYSIS POD FOR PLANAR VOIDS, NOISE-ADJUSTED LINEAR AREA-AMPLITUDE MODEL (TRANSDUCER NO. 2)

8.2.2.3 Effective Reflectivity Analysis: Initial Method, Improved Model.

Iowa State University developed more sophisticated methods for modeling the response of reflectors. For example, FBHs that allow the replacement of the linear area-amplitude approximation by a model that accounts for much of the nonlinearity observed in figures 30 and 34. The results of fitting the output from one of these models [30] to the experimental FBH data is shown in figure 40 (see appendix A for more details). The fit is clearly improved, although the model, which in contrast to the previous analysis (8.2.2.2), assumes there is no negligible or additive noise and appears to under-represent the amplitudes of the larger FBHs. Minitab does not provide a means for quantifying the improvement, since the new model is not available in closed form. Figure 40 also shows three regression lines that were described in sections 8.2.1, 8.2.2.1, and 8.2.2.2.

The ISU model results, in values of R_e shown in figure 41, are still not described well by the normal or lognormal distributions, despite the fact that the fit is better than the three linear regression lines (i.e., the hypothesis of normality is accepted at 0.02 significance by the Anderson-Darling test). Therefore, there is little difference between the fit of the normal and lognormal distributions.

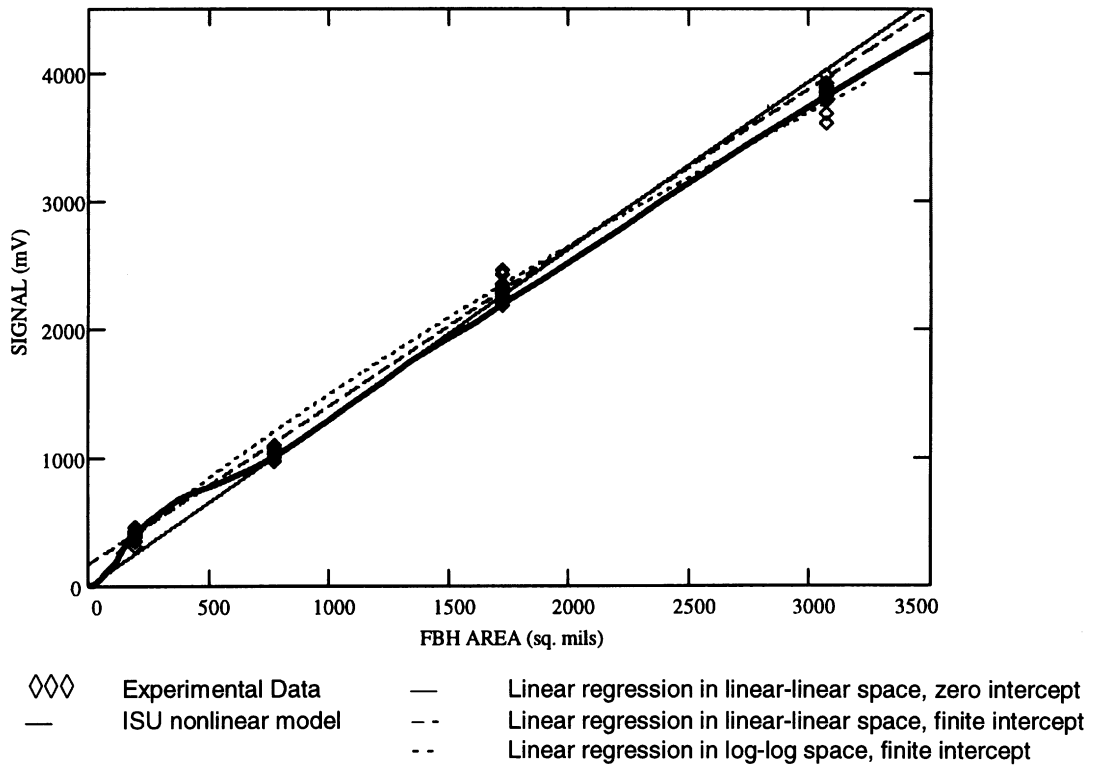


FIGURE 40. COMPARISON OF ISU MODEL WITH \hat{a} AND R_e REGRESSION LINES

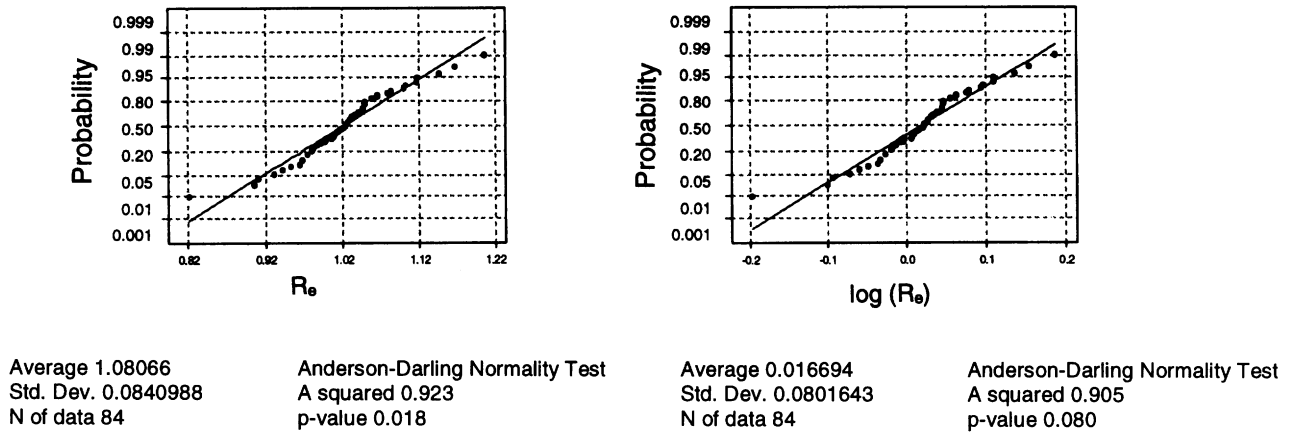


FIGURE 41. CUMULATIVE NORMAL PROBABILITY PLOTS FOR THE ISU MODEL (TRANSDUCER NO. 2)

Figure 42 shows the POD curves generated by applying Effective Reflectivity concepts to the ISU model (with location 1.0207 and scale 0.0611).

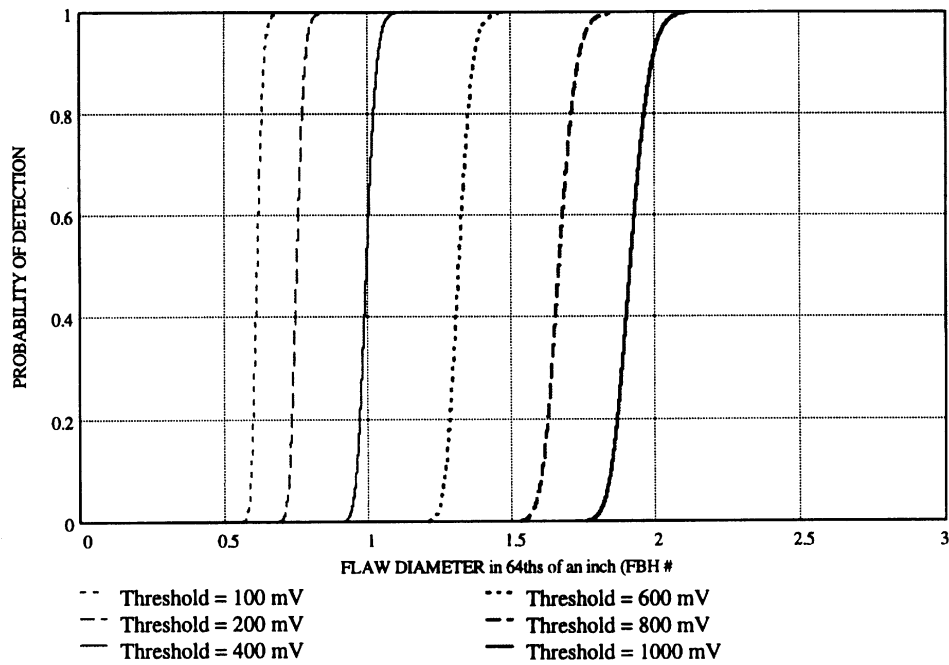


FIGURE 42. GE EFFECTIVE REFLECTIVITY ANALYSIS POD FOR PLANAR VOIDS, ISU NONLINEAR AREA-AMPLITUDE MODEL (TRANSDUCER NO. 2)

8.2.2.4 Effective Reflectivity Analysis: Revised Method.

The original Effective Reflectivity method was developed to be used with data from production inspections. This ensures that realistic flaw properties were used to consider the dependence of the signal on flaw size. Only flaws detected during real part inspections were used in the POD database. The revised Effective Reflectivity method makes up for the potential bias in the database, which occurs as a result of possible under-representation of low-detectability flaws. The new method [31 and 32] hypothesizes that the signal amplitudes (and the related R_e values) will fit the LEV distribution. The probability was calculated for obtaining various values of location and scale. These values were selected to maximize the overall probabilities for the entire dataset.

Although there was an improved confidence in the results, it lead to an increased complexity of the analysis. The iterative methods of solution necessary to identify the optimum values for the location and scale can be quite time consuming. Furthermore, the properties of the LEV distribution are less widely documented than those of the normal or lognormal distributions. Therefore, the number of tests available, which conform to the model assumptions, is more limited.

The results of analyzing the FBH data for transducer no. 2, using the revised Effective Reflectivity method and the ISU nonlinear area-amplitude model illustrated in figure 40, are shown in figure 43. Despite the conservatism that was believed to be built into this revised

method, comparisons with earlier POD plots show that the data in figure 43 results in a surprisingly high POD. This reflects one difficulty of incorporating nonlinear flaw model into the analysis. Figure 43 is based on using the model response of a no. 1 FBH to represent the calibration described in step a of section 8.2.2.1. As figure 40 suggests, different results are obtained when the calibration is based on the model for one of the larger FBHs (or on some other assumption).

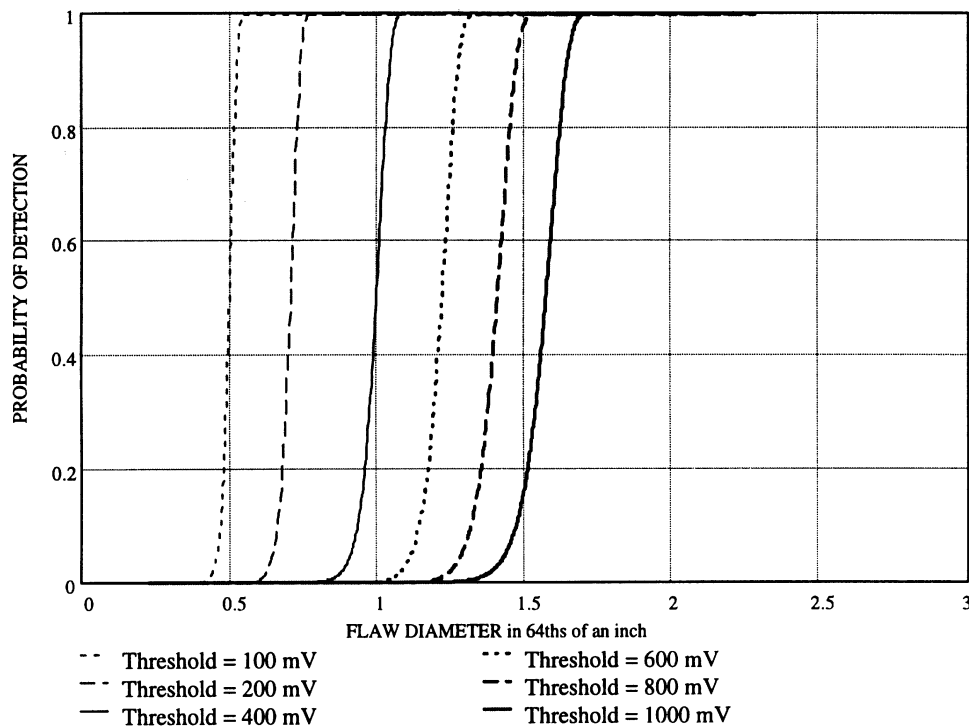


FIGURE 43. REVISED GE EFFECTIVE REFLECTIVITY ANALYSIS POD FOR PLANAR VOIDS, ISU NONLINEAR AREA-AMPLITUDE MODEL (TRANSDUCER NO. 2)

8.3 CONCLUSIONS.

8.3.1 Comparison With New Methodology Results.

Results from the new methodology can most easily be compared to those from the ASNT procedure by comparing the POD for a 400-mV threshold shown in figure 16 with the 50% confidence sizes listed in table 2. Results from the PF procedure (see figure 28) are not directly comparable with those of the new methodology, due to the limited range of thresholds for which the PF analysis could be completed. Comparing figures 33, 36, 39, 42, and 43 with figure 16 shows the actual POD values obtained depend on the method selected. For example, the POD for a no. 1 FBH, using a 400-mV threshold, varies from about 0.25 (R_e , figure 36) to about 0.75 (new methodology, figure 16). Other POD values are:

- 0.35 (A-hat, figure 33)
- 0.50 (ISU model, modified R_e , figure 43)

- 0.60 (ISU model, original R_e , figure 42)
- 0.70 (noise-adjusted R_e , figure 39).

Looking at the same data from another perspective, table 5 compares the hole sizes that exhibit a 50% POD as a function of threshold. The differences are most significant for small thresholds, corresponding to hole sizes out of the range in which data were available. Since fundamentally different assumptions are made by the models in this region, these differences are not surprising. The new methodology results in plausible POD (a) curves, yet the observed differences warrant further study.

TABLE 5. HOLE SIZE EXHIBITING 50% POD FOR SEVERAL METHODOLOGIES AS A FUNCTION OF THRESHOLD

Methodology	Threshold		
	200 mV	400 mV	800 mV
New Methodology	0.32	0.88	1.55
A-hat	0.67	1.02	1.55
Original R_e	0.85	1.18	1.68
Original R_e (noise adjusted)	0.35	0.98	1.63
Original R_e (ISU model)	0.75	1.00	1.63
Revised R_e (ISU model)	0.70	1.00	1.39

8.3.2 Typical Sets of Ultrasonic Detectability Data.

As noted at the beginning of section 8, the FBH data that are part of the current study are unusually well-behaved, in the sense that the scatter is not too large and, hence, they can be easily fit to a curve such as a regression line. By comparison, data from the inspections of natural flaws are more widely scattered and fewer in number. These differences make some of the methods considered above relatively inappropriate or difficult to use.

9. PHYSICS-BASED DESCRIPTIONS OF NOISE AND SIGNAL-PLUS-NOISE DISTRIBUTIONS: THEORY, EXPERIMENTAL VERIFICATION, AND STRATEGIES FOR INCORPORATION IN THE NEW METHODOLOGY.

9.1 MOTIVATION.

The first implementation of the new methodology, described in section 5.2, is based on an empirical determination of (1) the noise distribution and (2) the microstructure-induced variations in the flaw response to the physics-based, microstructure-free theoretical expectations. In order to determine the latter, a sample needs to be prepared with the microstructure of interest, containing a significant number (8 and 16 were used in these studies) of nominally identical reflectors. Experimentally comparing the responses of these reflectors determines the statistical model for the residuals, $\hat{\epsilon}$, describing the microstructural contributions to the flaw response variability. Additional sources of variability, such as the effects of scan plan and gate width, are also derived from the physics-based models, as were illustrated in section 7.2. As seen in

sections 7 and 8, this approach has allowed the POD predictions of FBHs that are in reasonable agreement with those of existing methodologies and which appear to provide stable answers for a number of conditions in which the existing methodologies experience difficulties. Moreover, for SHAs, predictions are made by the new methodology when none are delivered by existing methodologies. Thus, the use of physics-based models for the microstructure-free flaw response, combined with empirical measurements of noise distributions and microstructure-induced flaw response variations, have provided a substantial step forward in the author's ability to determine POD for the ultrasonic detection of interior flaws.

In the course of these investigations, it became clear that further improvements could be made by incorporating information from physics-based models about the form of the noise distribution and the microstructure contributions to the flaw response distribution. This section summarizes the information that was gained about the form of these distributions based on cooperative efforts of the POD and Fundamental Studies Tasks of the ETC. The section concludes with a discussion of how this information might be implemented in a second generation methodology, which will be developed in the future.

It should be noted that the studies reported in sections 9.2 and 9.3 were completed after the initial formulation of the methodology, which included the assumption that the noise followed a lognormal distribution. This work can thus be considered as a basis for refining that assumption in future work.

9.2 EMPIRICAL STUDIES OF NOISE DISTRIBUTIONS.

As the preceding discussion demonstrated, the estimation of underlying noise distributions is important for the estimation of PFA. Knowledge of this distribution is also important for the utilization of digital data, including such potential techniques as dynamic thresholding [33] and signal-to-noise, ratio-based material acceptance criteria [34]. This section covers an empirical analysis of ultrasonic data that was done to estimate the appropriateness and parameters of several closed-form statistical distributions. It will be followed (section 9.3) by a further discussion of noise distributions based on physics-based modeling results.

9.2.1 Mathematical Techniques.

Several statistical distributions were considered as possible candidates for describing the noise distribution. First, because of its simplicity, is the normal distribution with probability density function (PDF) given by

$$P_n(x) = \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(x-\mu_n)^2}{2\sigma_n^2}} \quad (16)$$

where μ_n and σ_n are the parameters to be determined. Second, natural extension of equation 16 is the lognormal distribution. Here, the PDF is

$$p_{\ln}(x) = \frac{1}{\sqrt{2\pi}\sigma_{\ln}x} e^{-\frac{(\log(x)-\mu_{\ln})^2}{2\sigma_{\ln}^2}} \quad (17)$$

with μ_{\ln} and σ_{\ln} defining the distribution. The third distribution is the LEV distribution. This distribution was selected due to its relation to the C-scan image formation process [35]. The LEV PDF is given by

$$p_{lev}(x) = \frac{1}{\sigma_{lev}} e^{-\frac{(x-\mu_{lev})}{\sigma_{lev}}} e^{-e^{-\frac{(x-\mu_{lev})}{\sigma_{lev}}}} \quad (18)$$

where μ_{lev} and σ_{lev} once again are the parameters to estimate. For completeness, another non-symmetric distribution was selected. This is the two-parameter Weibull distribution. Its PDF is

$$p_w(x) = \frac{\alpha}{\beta} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha} \quad (19)$$

with α and β being the unknown parameters.

In all cases, the parameters were estimated using the maximum likelihood method, described as follows [36]:

Suppose a set of random variables, X_1, X_2, \dots, X_n , was described by the probability distribution function $p(x; \Theta_1, \dots, \Theta_m)$ where the parameters $\Theta_1, \dots, \Theta_m$ have unknown values, take on the observed values x_1, x_2, \dots, x_n . Then the probability that n samples from the distribution take on these values is given by $L(x_1, x_2, \dots, x_n; \Theta_1, \dots, \Theta_m)$. This is called the likelihood function and, because each sample taken is independent, this likelihood can be expressed by

$$L(x_1, x_2, \dots, x_n; \Theta_1, \dots, \Theta_m) = p(x_1; \Theta_1, \dots, \Theta_m) \times p(x_2; \Theta_1, \dots, \Theta_m) \times \dots \times p(x_n; \Theta_1, \dots, \Theta_m) \quad (20)$$

If the authors take the likelihood to be a function of the parameters $\Theta_1, \dots, \Theta_m$, then maximizing the likelihood gives the parameter values for which the observed sample is most likely to have been generated. This set of parameters is called the maximum likelihood estimators. This operation can be done most easily by taking the natural logarithm of the likelihood

$$\ln(L(x_1, x_2, \dots, x_n; \Theta_1, \dots, \Theta_m)) = \sum_{i=1}^n \ln(p(x_i; \Theta_1, \dots, \Theta_m)) \quad (21)$$

and solving the set of equations, $j=1, 2, \dots, m$

$$\frac{\partial \ln(L(x_1, x_2, \dots, x_n; \Theta_1, \dots, \Theta_m))}{\partial \Theta_j} = 0 \quad (22)$$

for the values of the parameters will yield the maximum likelihood estimators. In practice, this was performed using MathCAD Plus Version 5.0 on a personal computer.

Once the maximum likelihood estimators are determined, the estimated PDF can be plotted on the same scale with the experimental PDF. This is done by plotting a histogram of the data set and then scaling it by the total number of points in the data set. The goodness of fit of the estimated PDF can then be determined by visual inspection of these plots. This is a very subjective method; however, several quantitative comparison techniques were also identified. The first was the root mean squared (rms) error between the experimental and estimated PDF. Both weighted and unweighted rms errors were used. The Wirsching-Carlson W and S statistics

$$W^2 = \frac{1}{n} \sum_{i=1}^n (p_{\text{exp}}(i) - p_{\text{est}}(i))^2 \quad (23)$$

$$S = \frac{1}{n} \sum_{i=1}^n |p_{\text{exp}}(i) - p_{\text{est}}(i)| \quad (24)$$

are other quantitative or objective evaluation methods which were used [37]. For the data sets considered for this work, the fits were sufficiently good that the probability plots did not differentiate the techniques and, in addition, the results from all the quantitative comparisons were consistent, so in the interest of simplicity, only the rms error values were used. The probability plots did, however, present useful insight into the areas where the quality of the fit was not as good. As in most situations, this proved to be in the low probability or tail area of the distribution.

9.2.2 Experimental Techniques.

An important part of doing this analysis is choosing a data acquisition method which guarantees consistent and statistically meaningful results. The main parameters identified which could affect the statistical quality of the fit were the number of data points used to form the estimate as well as the spacing of the pixels relative to the beam diameter of the transducer. As the paper that resulted from this work [38] discusses in more detail, the experimental and estimated PDF were calculated for the different data sets and the resulting rms errors plotted. The rms errors showed little change in the range from 0.008" to 0.064" pixels and then vary wildly for the 0.128" pixels. The rms error also increases as the amount of material sampled is reduced, until it reaches a point where it behaves inconsistently. Based on these results and to produce a consistent analysis, it was decided that all data would be acquired over as large an area of the sample block as possible and must contain a minimum of 5,000 pixels regardless of the transducer beam diameter.

Ultrasonic test data were taken with many combinations of data acquisition parameters. The parameters consisted of a C-scan gate duration, transducer frequency, transducer beam

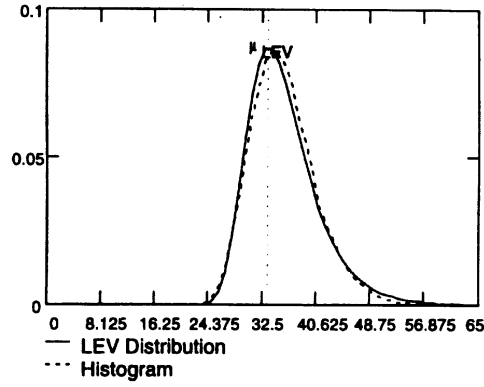
diameter, and the material microstructures. To be an effective method of characterizing grain noise, the estimation technique should be robust under a variety of data acquisition conditions. The effect of each of the first three conditions on the estimation results was individually addressed by taking data from three identically designed sample blocks cut from three different Ti forgings. The fourth condition was addressed implicitly by taking the data from the three different sample blocks.

To test the effect of each of these conditions, the following was done:

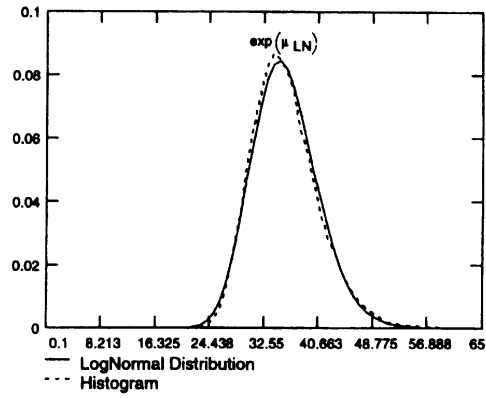
- C-scan gate duration—Data were taken from a 5-MHz transducer with a 0.100" focal plane beam diameter from the sample blocks with 1-, 2-, 4-, and 8- μ s gate widths, all centered around the focal plane of the transducer, which was placed at 1" metal travel, i.e., 1" below the metal surface. The estimates and associated rms errors were calculated and the results plotted.
- Frequency—Data were taken with 5-, 7.5-, and 10-MHz transducers, all with a nominal 0.100" beam diameter, on the sample blocks using an 8- μ s gate width centered around the focal plane of the transducer, which was placed at 1" metal travel. The estimates and associated rms errors were calculated and the results plotted.
- Beam diameter—Data were taken with 0.069", 0.100", and 0.116" focal plane beam diameter transducers, all with a 5-MHz nominal center frequency, on the sample blocks using an 8- μ s gate width centered at the depth of focus of the transducer, which was placed at 1" below the entry surface. The estimates and associated rms errors were calculated and the results plotted.

9.2.3 Results.

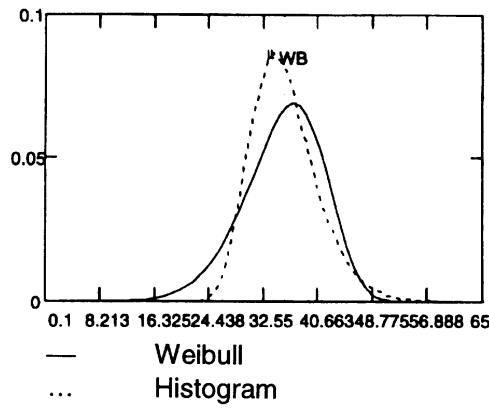
Plotting the results of these experiments demonstrated the skewed nature of C-scan noise distributions, as shown in figure 44. As a result, the lognormal and LEV estimates give much better fits than those produced by the normal and Weibull distributions. They also show that both the lognormal and the LEV estimates are robust with respect to changes in data acquisition parameters. No clear trends could be found in the data with respect to any of the four above defined conditions. Upon looking at the data, there were no clear distinctions in performance between the lognormal and LEV estimates. This would lead to the conclusion that either statistical distribution would suffice, and the selection should be controlled by other factors such as computational intensity.



(a)



(b)



(c)

FIGURE 44. PLOTS OF ESTIMATED AND EXPERIMENTAL DISTRIBUTIONS FOR A 5-MHz, 0.100" DIAMETER TRANSDUCER ON A Ti-17 TEST BLOCK (a) LEV FIT WITH $\mu = 33.405$, $\sigma = 4.221$, $\text{rms} = 0.003$, (b) LOG NORM FIT WITH $\mu = 3.567$, $\sigma = 0.135$, $\text{rms} = 0.002$, and (c) WEIBULL FIT WITH $\mu = 35.728$, $\sigma = 4.931$, $\text{rms} = 0.009$

Two potential applications of these distributions, the determination of a SNR (signal-to-noise ratio)-based reject criterion and the calculation of the probability of false alarm, are highly dependent on the low probability or tail area of the distribution. Looking at the tail area of the distributions considered here (see figure 45) reveals that the estimate from the LEV distribution lies above the experimental distribution and the estimate from the lognormal below. These results were typical for the data considered here. Based on this trend, one of the estimated distributions may be better suited for a particular application than the other.

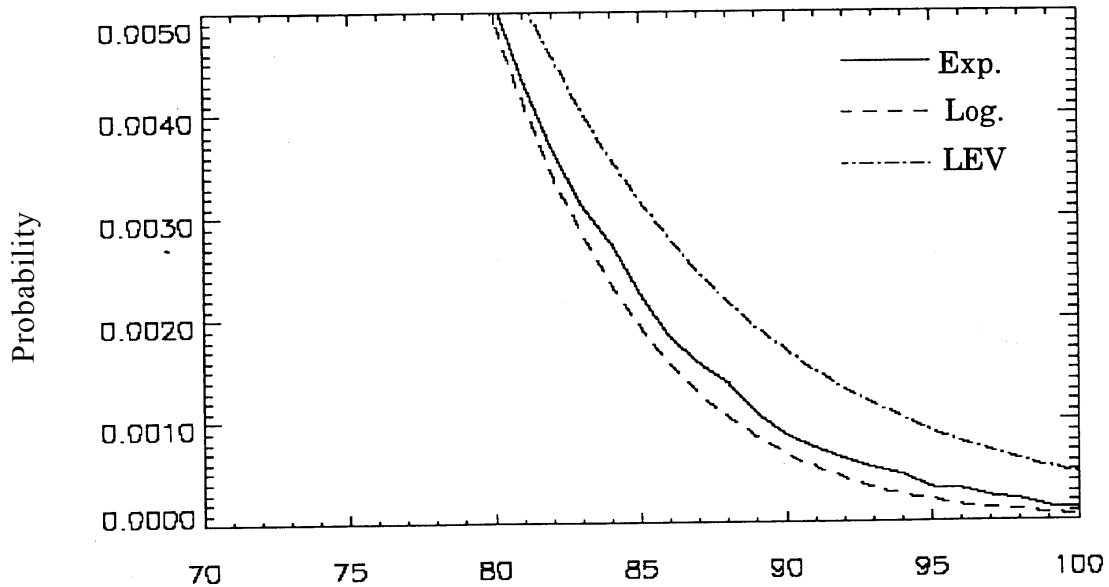


FIGURE 45. THE LOW PROBABILITY OR TAIL AREA OF THE EXPERIMENTAL AND ESTIMATED PDFs DERIVED FROM GRAIN NOISE DATA COLLECTED WITH A 5-MHz, 0.100" BEAM DIAMETER TRANSDUCER ON A Ti-17 SAMPLE BLOCK

For example, when the estimated distributions are used to calculate the noise term in a SNR calculation, the noise term derived using the lognormal curve would necessarily be lower than that derived using the LEV estimate. With all other terms in the SNR calculation being equal, this would result in the SNR calculated from the lognormal distribution being higher than that from the LEV. Using a material acceptance criterion based on a constant SNR threshold, the estimate derived using the lognormal distribution would necessarily reject flaws with lower amplitude signals.

Next, consider using the estimate to calculate PFA. The PFA is defined as the amount of area under the PDF of the noise, which lies above the detection threshold value. Therefore, a PFA calculated using the LEV estimate would be larger than that calculated using the lognormal estimate. When trying to evaluate the practicality of an inspection technique, it is generally better to overestimate the PFA. Thus, when using this technique in practice, it is important to consider the application as well as the magnitude of quantitative fit measurements.

9.2.4 Conclusions.

This work discusses an experimental method of estimating the parameters of a closed-form statistical distribution from ultrasonic C-scan data. Using this method, it was shown that reasonable estimates of the Ti grain noise statistical distribution can be obtained using either the lognormal or largest extreme value (LEV) distribution. Both of these distributions provided estimates that had low rms errors with respect to the experimentally determined distributions under a wide variety of ultrasonic scan conditions. Although generally acceptable, these estimates exhibit deficiencies in the tail of the distribution that could cause concern in certain applications. This provides a motivation for the further studies reported in section 9.3.

The results discussed here were only for forged Ti microstructures. Preliminary results indicate that this method works equally well on data from billet microstructures. However, this needs to be more closely investigated in the future.

9.3 MODEL-BASED STUDIES OF NOISE DISTRIBUTIONS.

The noise deficiencies of the lognormal or LEV distributions in fitting the tail of the gated peak-to-peak noise distribution motivated further studies as to the proper functional form of the gated peak-to-peak noise distribution. The lognormal distribution was used in the initial implementation of the methodology as discussed in section 5.2. However, neither the lognormal nor LEV distributions were able to satisfactorily describe the behavior of the high-amplitude tail of the distribution. This lack of goodness of fit is a concern in the estimation of PFA for two reasons: (1) it is a matter of some economic concern to the original equipment manufacturers (OEMs) and (2) in the implementation of improved detection techniques, such as noise-based thresholding and signal-to-noise ratio-based rejection criteria. Thus, further examination of the problem was conducted, leading to improved fits, as described in this section. The essential idea of these developments is that the noise within a gate is sampled independently in different time subintervals within that gate. Hence, the gated peak-to-peak noise distribution is better described as the maximum of multiple random variables (RVs) rather than a single RV. Although resulting descriptions of the noise distribution have not been implemented in the current methodologies, they are available for the next generation.

9.3.1 Theoretical Models for Gated Peak-to-Peak Noise.

As indicated in section 5.1, estimating the underlying noise distributions is very important in a signal-to-noise ratio-based approach, estimating the POD, PFA, and the risk determination via the ROC method. This subsection covers a theoretical development of the modeling and estimation of ultrasonic grain noise in large-grained alloys. Analysis based on these developments is reported in section 9.3.2, where data from forged and billet microstructures of titanium alloys are used to verify the model.

Possibly the most straightforward approach to modeling the noise distribution of gated peak-to-peak signals is the extreme value theory [39 and 40]. This is an asymptotic approach that produces three possible distributions for the maximum signal occurring in the gate. The LEV distribution occurs when certain conditions are held. The continuous signal in the gate (A-scan,

rectified A-scan, etc.) is comprised of a number of discrete instantaneous signals. It is assumed that each discrete value in the gate is the result of observing a random variable, which are all independent and identically distributed (IID). (The term “copies” will be used in the remainder of this section to describe this situation.) One thinks of the total gate as being divided into a number (N) of subregions, in which the same distribution function is copied. When the RVs are IID, the maximum signal in the gate is determined by taking the maximum value of N samples taken from this copied distribution. Allowing the number of possible discrete, instantaneous observations N to approach infinity, leads to the N LEV distribution, i.e., when the discrete RV has a Rayleigh distribution. Thus, if each discrete signal is an envelope amplitude from a complex normal process, then the limiting or asymptotic noise distribution is the LEV. This modeling approach was first applied to make enhancements to the effective reflective method [31, 35, and 41], as discussed in section 9.2.

Another approach [42] has also been developed, based on discrete gate RVs, that is less restrictive in its mathematical assumptions. In this setup, the individual RVs do not necessarily have all the same distribution, and/or only a finite number of discrete instantaneous signals are employed in modeling the noise distribution of the gated peak-to-peak signal. Due to the nature of this development, it is possible to relate the number of individual RVs in the gate to parameters in the inspection process. In the context of ultrasonic inspection, the envelope amplitude may be thought of as a continuous function that connects the peaks of the individual cycles of the RV waveform. The waveform quantity is sometimes represented by the magnitude of the analytic signal [43] and is closely related to the rectified output of typical ultrasonic instruments.

The discrete instantaneous RVs are considered to be envelope amplitudes because the output of the gate is the maximum peak-to-peak response. Various arguments indicate that this maximum peak-to-peak response can be approximated by twice the maximum of the independently sampled envelope amplitudes. The development of the Rayleigh distribution as the distribution of discrete RVs is analogous to developments of the signal amplitude distribution in signal detection theory [43]. When a sample of material is inspected with an interrogating beam, some of the energy is absorbed and some is reflected. The reflected, or backscattered, signal that is observed at any instant (depth) is a function of the resultant of individual energy contributions from each of many reflectors (grains, microstructure anomalies, etc.) in the ensounded material. This can be described by a random walk in the complex plane or the random flights model [44-48] (cf. figure 46). If the real and imaginary parts of the reflector signals of the random walk are IID with independent normal distributions, then the resultant envelope amplitude follows a Rayleigh distribution. The same conclusion holds (under the IID assumption) by virtue of the Central Limit Theorem if the number of reflectors is large, and this is the argument generally used to develop the distribution of an individual RV in the gate.

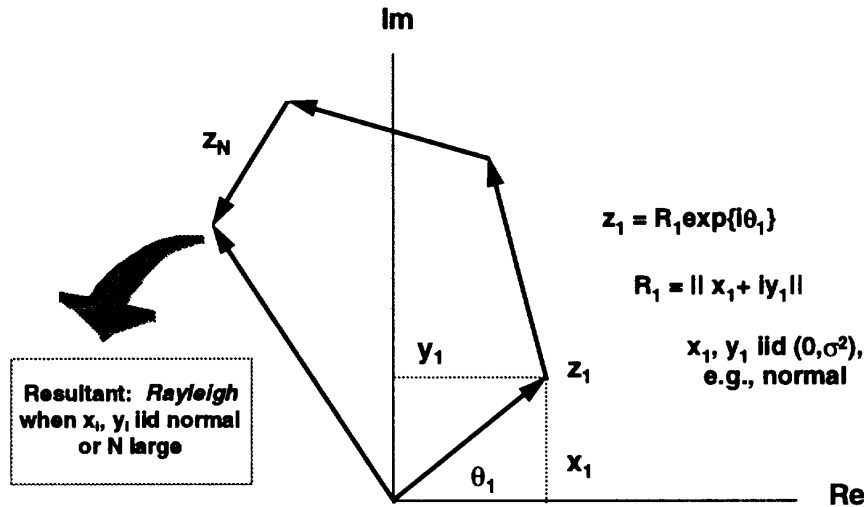


FIGURE 46. RANDOM FLIGHTS MODEL

If the grains that produce an instantaneous gate signal are not homogenous in reflecting properties and/or only a small number of scatterers are ensonified, the limiting Rayleigh distribution may not hold. Therefore, the Rayleigh distribution may not be applicable to the discrete RVs of the gate signal. Thus, it would be of interest to have other candidate distributions for study. Three other distributions that might be used are now considered.

- The K distribution [49 and 50] is a solution to the two-dimensional random walk problem in the complex plane. It generalizes the standard normal assumptions in that, when the amplitude of each reflector has a K distribution, the resultant amplitude also has a K distribution, i.e., the K distribution is reproductive. The use of the K distribution as a generalization of the Rayleigh distribution has been previously considered [51] due to its ability to represent both the normal distribution situations and the non-normal distribution situations.
- The original development of the effective reflectivity method [8] empirical studies led to the lognormal distribution, so this distribution could describe the gated-peak-to-peak noise distribution. However, it is interesting to note that the original application of the effective reflectivity method was for small-grained, powder alloys and the distribution was for the gated-peak-to-peak normalized flaw response (i.e., R_e) (and not an individual RV).
- The Weibull distribution is one of the smallest extreme value distributions and, thus, might not be warranted as a candidate. However, when the shape parameter of the Weibull distribution has the value 2, the Weibull distribution reduces to the Rayleigh distribution, so that the Weibull could also be studied.

To summarize, there are five natural candidate distributions for use in the study of the gated peak-to-peak noise: LEV, Rayleigh, K, lognormal, and Weibull. However, the LEV and lognormal might only represent the gated peak-to-peak noise and not an individual RVs

producing the gated output noise. It will be asserted in section 9.3.2.3 that the asymptotic distribution of the maximum of each of these five distributions is the LEV. Thus, if the number N of independent copies in the gate is large, the use of the LEV distribution can be based on a number of underlying distributions that have merit. Moreover, earlier studies [38] of ultrasonic grain noise have shown that LEV overestimates the right tail of observed noise distributions in certain cases, while lognormal underestimates the right tail (cf. figure 45). Considering these two distributions, as associated with a finite number of individual RVs in the gate like the Rayleigh, K, and Weibull might improve upon these observed results, consistent with the physics-based model presented in reference 42.

Each of the five distributions will be used as the distribution of a discrete RV gate in an estimation scheme where the number of discrete instantaneous signals is finite and unknown, i.e., must be estimated. Since all distributions are known in form, the method of maximum likelihood estimation (MLE) will be employed. Use of MLE is well established. However, in the present situation, using MLE led to some interesting challenges which are discussed. It should be noted that the study only focused on complete distributions. However, the major interest is in the right tail of the noise distribution. Future efforts will consider various means of weighting the observations so more emphasis is placed on accurately estimating the right tail of the noise distribution. This could also improve the use of the LEV and/or lognormal distributions as gated peak-to-peak noise distributions [52].

9.3.2 Empirical Investigation of Gated Peak-to-Peak Noise.

In this section, the analysis procedure is outlined and the investigation results are presented for each of the five candidate distributions discussed in section 9.3.1 (LEV, Rayleigh, K, lognormal, and Weibull) for the distribution of a discrete gate RV. An MLE procedure was used to obtain parameter estimates for the distribution of the maximum unknown number of IID copies in each candidate distribution. The fitted distributions for envelope amplitude were derived by substituting parameter estimates, including an estimated number of independent copies, into the form of the distribution of the maximum in each case. Both graphical and statistical goodness-of-fit (GOF) assessments were made from the resulting distributions.

9.3.2.1 Maximum Likelihood Procedure (Preliminary).

The five natural candidate probability density functions (PRFPDF's) for discrete gate observations were:

$$\text{Lognormal} : f(y | \mu, \sigma) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln(y) - \mu)^2}{2\sigma^2}\right\}, \sigma > 0 \quad (25)$$

$$\text{Rayleigh} : f(y | \beta^2) = \frac{y}{\beta^2} \exp\left\{-\frac{y^2}{2\beta^2}\right\} \quad (26)$$

$$\text{Weibull : } f(y | a, b) = \frac{b}{a^b} y^{b-1} \exp\left\{-\frac{y^b}{a^b}\right\}, a, b > 0 \quad (27)$$

$$\text{LEV : } f(y | \mu, \sigma) = \frac{1}{\sigma} e^{\left\{\frac{-(y-\mu)}{\sigma}\right\}} \exp\left\{e^{\left\{\frac{-(y-\mu)}{\sigma}\right\}}\right\}, \sigma > 0 \quad (28)$$

$$\text{K : } f(y | b, m) = \frac{2b}{\Gamma(m)} \left(\frac{by}{2}\right)^m K_{m-1}(by), b, m > 0 \quad (29)$$

The literature contained discrepancies concerning the form of the K distribution [53], which the authors have attempted to resolve [50]. However, this distribution is quite flexible or may be generalized [54], to yield all candidate distributions under consideration for particular parameter value settings.

For an unknown number (N) of IID copies of any candidate distribution and n sample points (y_i) used in the procedure, the likelihood for the maximum peak-to-peak noise in the gate was taken to be [42]:

$$L(\theta, N | y_i, i = 1, 2, \dots, n) = \prod_{i=1}^n N[F(y_i | \theta)]^{N-1} f(y_i | \theta) \quad (30)$$

where θ generically represents each candidate distribution's parameter vector, and F is the cumulative distribution function (CDF). The actual MLE procedure used was a two-stage estimation scheme where the unknown number of copies in the gate from each candidate distribution, N, was estimated in stage one, then fixed, and the remaining parameters re-estimated in stage two, given the fixed value for N from stage one. No appreciable differences in parameter estimates resulted from stage one (free optimization) to stage two (conditional optimization).

As the time required to compute MLEs was approximately 5 hours for all five candidates (for $n \cong 15000$ on a Sun SPARCstation 10, using Splus software from MathSoft, Inc.), a modification to the procedure was made for two reasons: (1) expediency and (2) validity. A routine was built which subdivided the range of observations into bins, accumulating observations in bins with a minimum count of ten observations per bin. One pass was made from the left tail of the empirical distribution to the mode and another from the right tail to the mode. As this heuristic approach will typically leave holes in the range of coverage of the resulting subintervals, an aggregation of bins was made which left no coverage holes. The minimum number of observations per bin after aggregation was first taken to be 5 and then 2 with no appreciable differences in parameter estimation. The data, after binning, were summarized by the bin endpoints and the counts of observations per bin.

The likelihood associated with this binning procedure was derived from the multinomial distribution, with probabilities provided by the candidate distribution CDFs:

$$L(\theta, N|(c_i, l_i, r_i), i = 1, 2, \dots, k) = \prod_{i=1}^k \left[F(r_i|\theta)^N - F(l_i|\theta)^N \right]^{c_i}, \quad (31)$$

where (c_i, l_i, r_i) are the count, left-hand endpoint, and right-hand endpoint for the i -th of the k bins. Binning reduced the computation time by an order of magnitude (5 minutes for the cycle of five candidate distributions for $n \cong 15000$, $k \cong 50$).

9.3.2.2 Maximum Likelihood Procedure (Final).

It became evident after the first few attempts at fitting equation 31 to the data that some form of interval censoring had taken place in the recording of the envelope amplitude data. All observations in a C-scan dataset were either integer or even. This rendered equation 31 inappropriate when applied to continuous data. Moreover, the exact type of censoring, truncation (chopping), or rounding was not known, and thus a modification to equation 31, which allowed for either, was introduced:

$$L(\theta, N|(c_i, l_i, r_i), i = 1, 2, \dots, n) = \prod_{i=1}^k \left[F(r_i + \delta_1|\theta)^N - F(l_i + \delta_1 - \delta_2|\theta)^N \right]^{c_i} \quad (32)$$

where table 6 provides values of δ_1 and δ_2 :

TABLE 6. LIKELIHOOD MODIFICATIONS FOR CENSORING

	Integer Observations	Even Observations
Rounding	$\delta_1 = 0.49999999$ $\delta_2 = 1.00000001$	$\delta_1 = 0.49999999$ $\delta_2 = 2.00000001$
Truncation	$\delta_1 = 0.99999999$ $\delta_2 = 1.00000001$	$\delta_1 = 0.99999999$ $\delta_2 = 2.00000001$

Equation 31 is the likelihood upon which all conclusions are based.

9.3.2.3 Analytical Asymptotic Results.

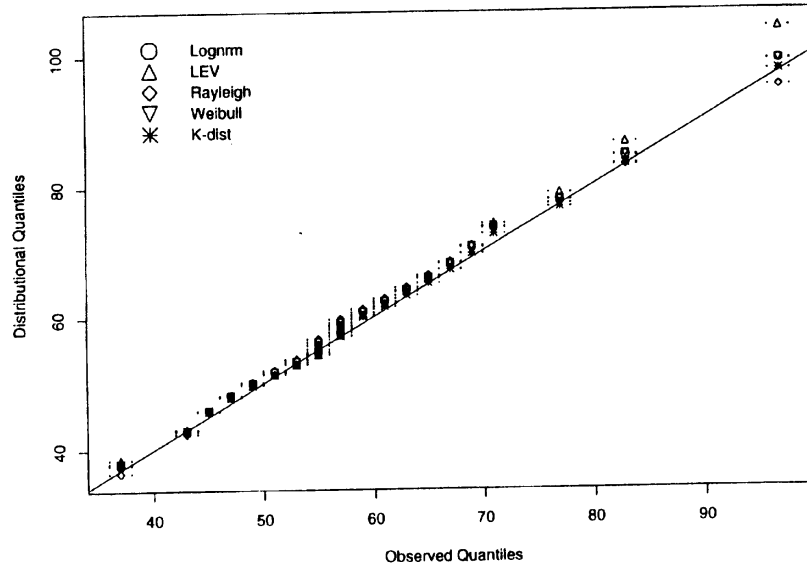
Each of the candidate distributions exhibits the LEV as the distribution of the maximum in the gate when the number of IID copies, N , approaches infinity. This result is well known for the lognormal and the normal distribution but is new for the Rayleigh, Weibull, and K distributions. Hence, the LEV distribution should produce an estimate of N near 1 and perform well in a GOF assessment if estimates of N are large for the other candidate distributions.

9.3.2.4 Goodness of Fit Assessments.

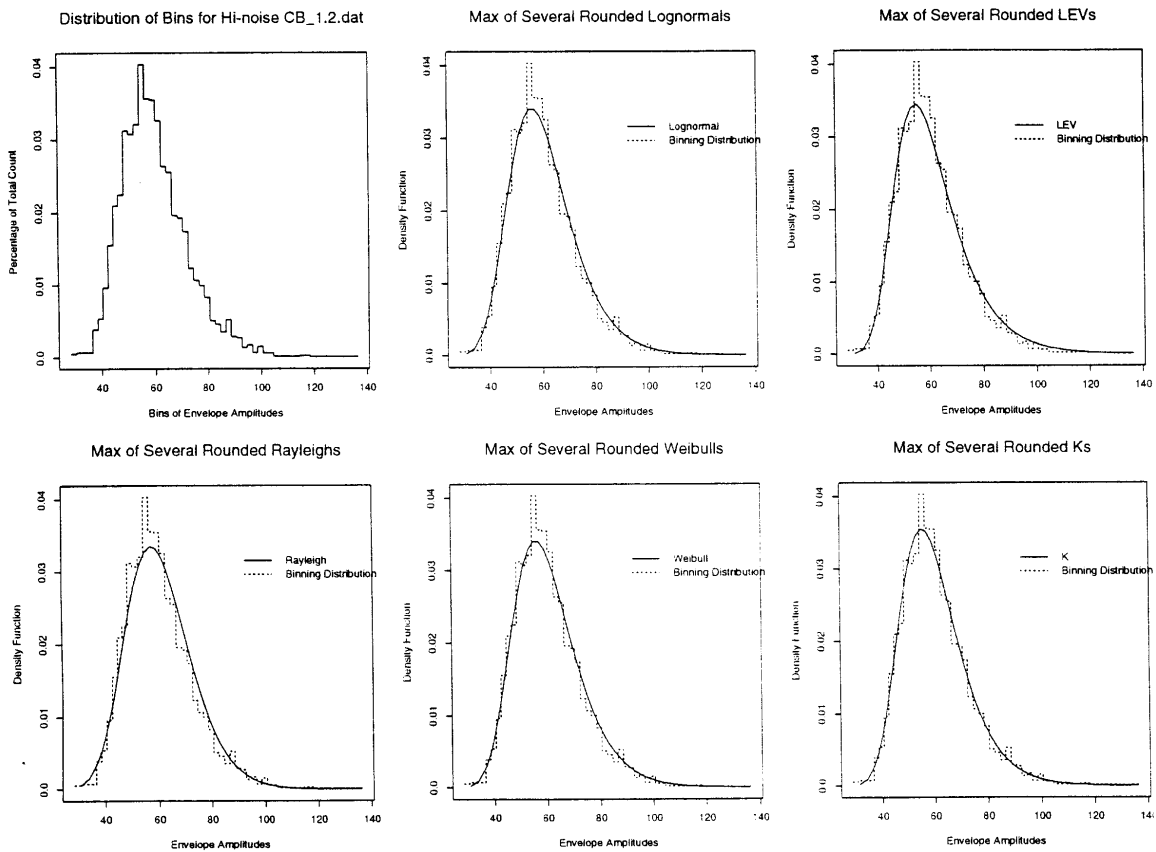
Seven billet data sets and twenty-four forging data sets were analyzed. They differed in overall noise level but did not differ significantly regarding the results they produced by type of specimen (see section 9.3.2.6). Both graphical and statistical procedures were used to assess the GOF for each of the fitted candidate distributions. Initially, B-splines were used to fit the empirical CDF, with midpoints of the data grid used as knots for the splines. An empirical PDF was defined as a derivative of the B-spline, which was fitted to the empirical CDF [55]. The Splus version of the spline derivative did not allow constrain for non-negativity in the empirical PDF. In other words, when the B-spline fit to the empirical CDF is made and differentiated to obtain the PDF, there is no assurance in the fitting procedure that the PDF will always be positive, as must be the case for any proper probability distribution function. Therefore, the assessment was discontinued, and a comparison was done between the empirical histogram with the fitted PDFs and Q-Q plots. Accompanying plots provide more detail. Figure 47 CB_1.2.dat provides results for one billet run made on the high-noise region of the contaminated billet (see section 10.2). Figure 48 S94D_4X.dat provides results for one forging run. The results presented in figures 47 and 48 show the assumptions of both data rounding (parts a and b) and data truncation (parts c and d). The “whiskers” in the Q-Q plots signify flat portions of the empirical CDF.

Three statistical measures were applied to the fitted (continuous) distributions in the first round of analyses: Kullback-Leibler divergence, Hellinger distance, and the Kolmogorov-Smirnov (K-S) test statistic. Upon realizing that the data were interval-censored, the K-S statistic was modified accordingly (dropped the Kullback-Leibler divergence and Hellinger distance, as these are continuous measures) and incorporated the χ^2 test statistic based on the bins into which the data were gathered. Tables 7 and 8 provide examples of the results for the billet data set and the forging data set (whose results are shown graphically in figures 47 and 48).

The values of the Chi-square and K-S statistics in tables 7 and 8 should be interpreted in the usual manner (large values indicate lack of statistical fit). The p-values are associated with the Chi-square statistic and should be used as a figure of merit only. A small p-value indicates significance and have corresponding large values of the Chi-square statistic. The data were binned for analysis purposes to just eliminate gaps in the tails. Thus, the number of bins can experience large changes from situation to situation, and there are cases in which the expected number of observations per cell is not at least 5. Also, the number of bins is large, tending to produce unwanted statistical significance. Given the nature of the Chi-square statistic as applied in tables 7 and 8, one would not use the ordinary guidelines in declaring statistical significance. Therefore, the Chi-square p-values should be used with caution. The K-S statistic is corrected for binning and does not suffer the problems that the Chi-square does. It is the more reliable of the two statistics. Unfortunately, simulation is required to obtain the p-values for the K-S statistic. This was not done in the Phase I effort. With these caveats, possibly the only safe conclusion is that the LEV does not fit the billet data. In this vein, note that, except for the LEV distribution, the situation is mixed between the two tables for the values of the K-S statistic.

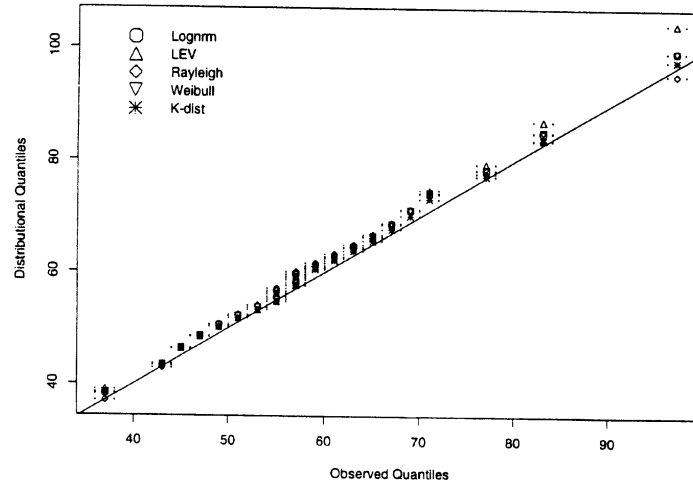


(a) Q-Q Plot Assuming Rounding in the Data

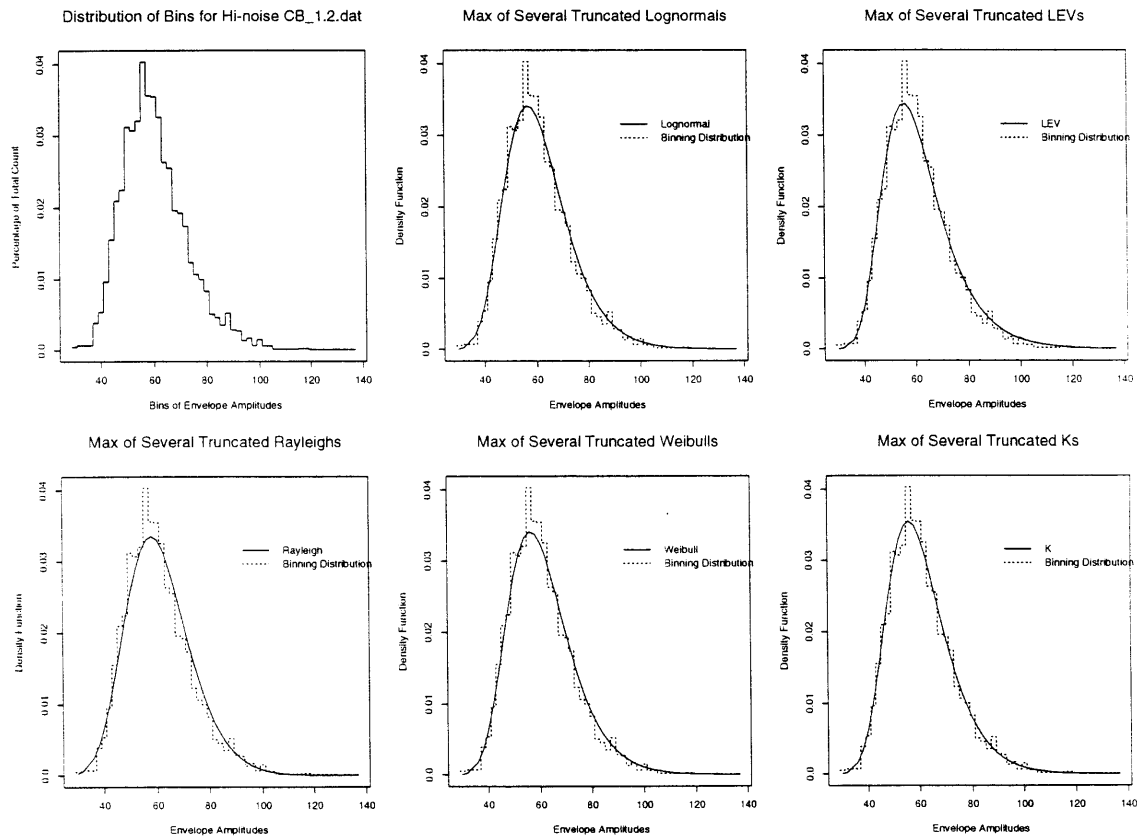


(b) Comparison of Several Distributions to the Noise Histogram, Assuming Rounding

FIGURE 47. FITTING VARIOUS DISTRIBUTIONS TO NOISE DATA FROM THE HIGH NOISE REGION OF THE CONTAMINATED BILLET

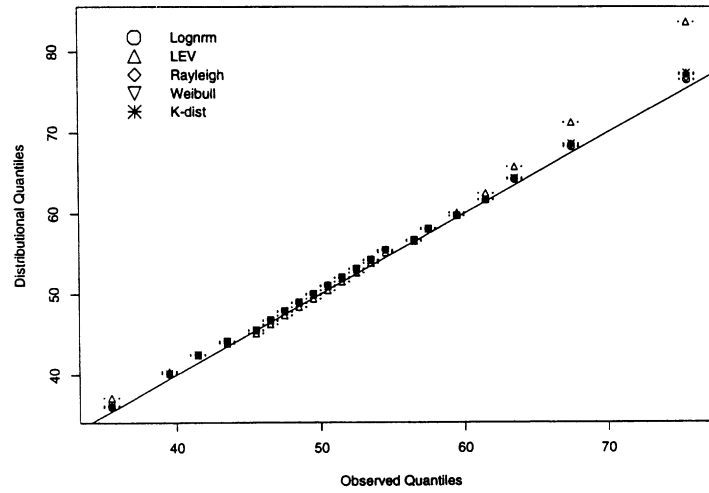


(c) Q-Q Plot Assuming Truncation in the Data

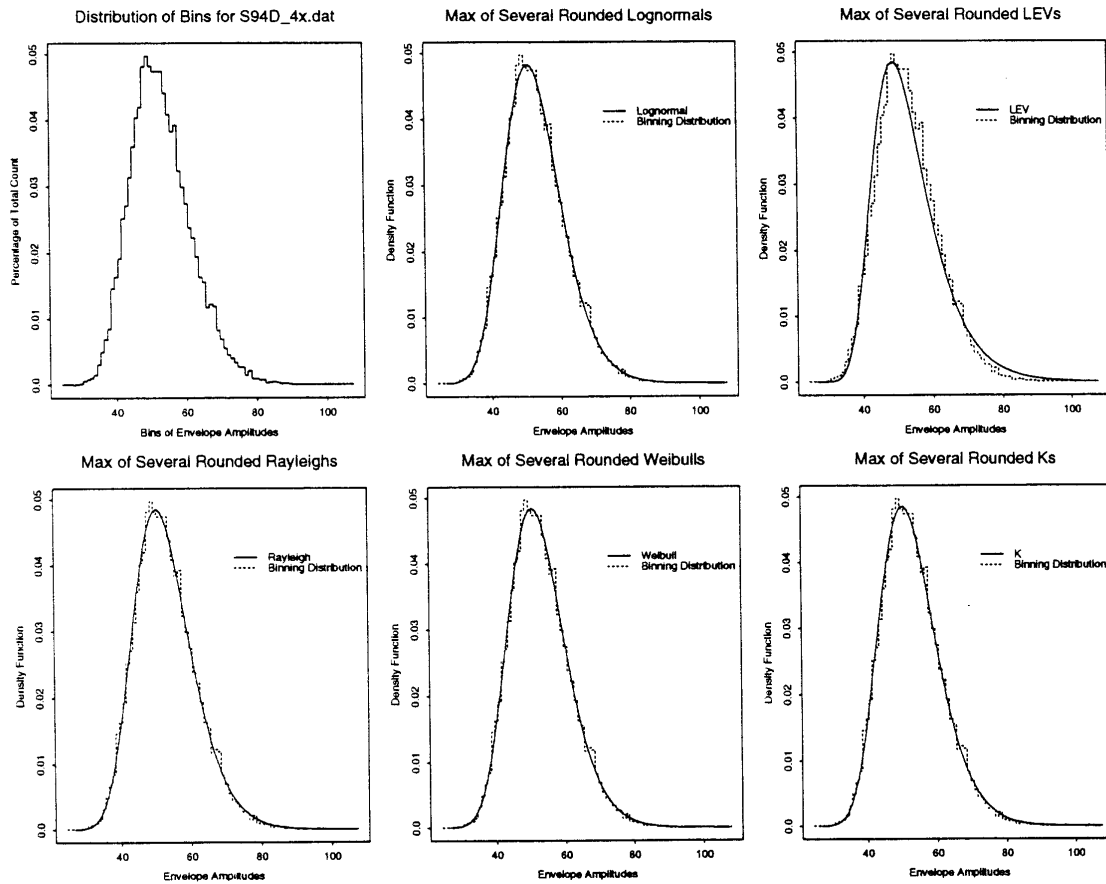


(d) Comparison of Several Distributions to the Noise Histogram, Assuming Truncation

FIGURE 47. FITTING VARIOUS DISTRIBUTIONS TO NOISE DATA FROM THE HIGH NOISE REGION OF THE CONTAMINATED BILLET (Continued)

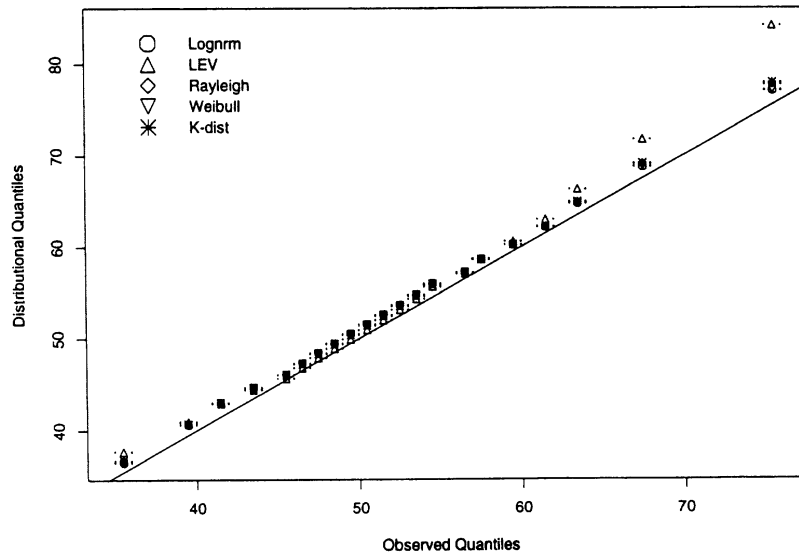


(a) Q-Q Plot Assuming Rounding in the Data

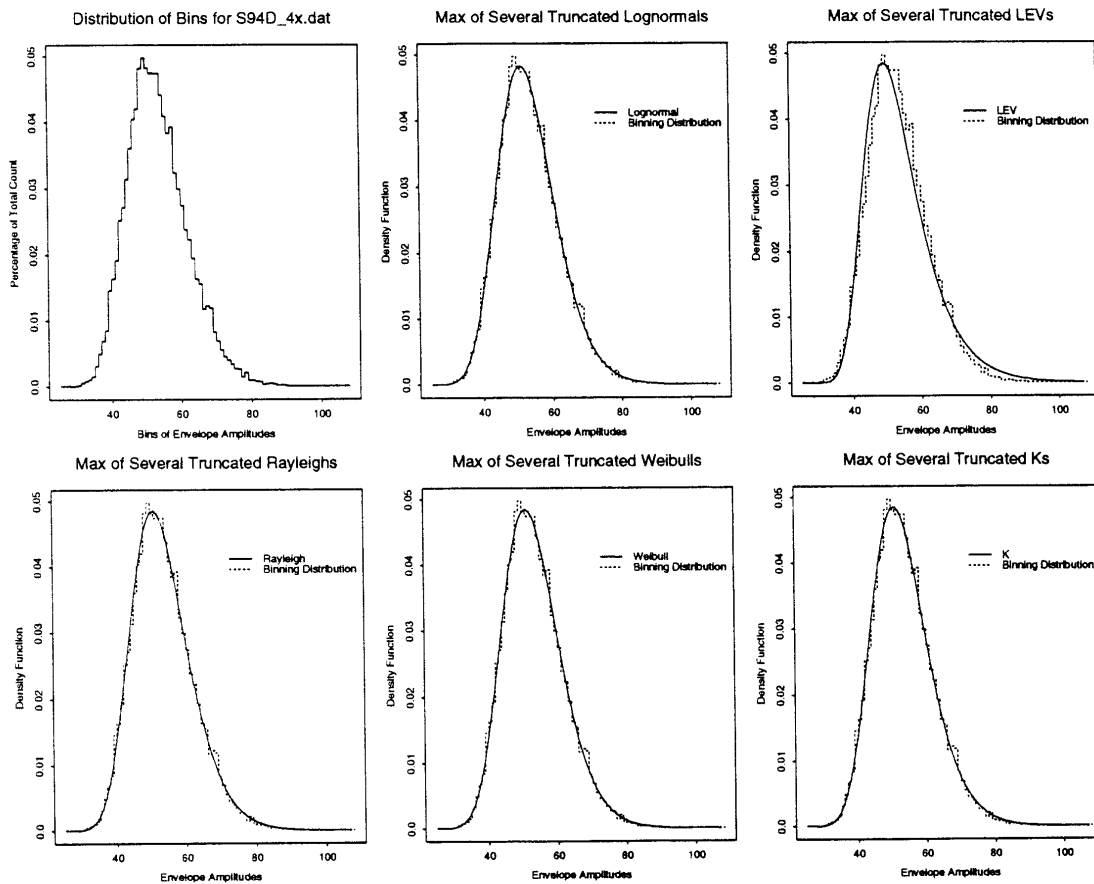


(b) Comparison of Several Distribution to the Noise Histogram, Assuming Rounding

FIGURE 48. FITTING VARIOUS DISTRIBUTIONS TO NOISE DATA FROM A FORGING



(c) Q-Q Plot Assuming Truncation in the Data



(d) Comparison of Several Distributions to the Noise Histogram, Assuming Truncation

FIGURE 48. FITTING VARIOUS DISTRIBUTIONS TO NOISE DATA FROM A FORGING
(Continued)

TABLE 7. STATISTICAL MEASURES FOR BILLET NOISE DATA

Hi-Noise CB_1.2.dat Rounded						
	param 1	param 2	N	Chi-sq	P-value	K-S
Lognormal	3.899	0.263	2.454	69.498	0.001	110.535
LEV	60.286	10.680	0.599	138.629	0.000	153.849
Rayleigh	25.474		10.367	159.524	0.000	77.082
Weibull	1.435	23.056	32.609	71.897	0.000	113.008
K	0.161	5.697	22.157	55.553	0.020	46.686
Hi-Noise CB_1.2.dat Truncated						
	param 1	param 2	N	Chi-sq	P-value	K-S
Lognormal	3.903	0.263	2.523	69.280	0.001	110.584
LEV	58.776	10.680	0.723	138.628	0.000	153.849
Rayleigh	25.566		10.648	157.596	0.000	76.919
Weibull	1.437	23.152	33.602	71.814	0.000	113.040
K	0.161	5.755	22.840	55.472	0.020	46.700

TABLE 8. STATISTICAL MEASURES FOR FORGING NOISE DATA OBTAINED WITH A TRANSDUCER HAVE A NOMINAL CENTER FREQUENCY OF 5 MHz AND USING A 4- μ sec TIME GATE

S94D_4x.dat Rounded						
	param 1	param 2	N	Chi-sq	P-value	K-S
Lognormal	3.911	0.176	1.332	56.724	0.698	65.404
LEV	52.316	7.576	0.624	1597.767	0.000	495.579
Rayleigh	19.625		22.840	63.488	0.495	115.235
Weibull	2.143	29.934	17.974	59.087	0.617	69.570
K	0.951	171.624	23.662	33.532	0.999	143.172
S94D_4x.dat Truncated						
	param 1	param 2	N	Chi-sq	P-value	K-S
Lognormal	3.918	0.175	1.363	56.684	0.699	65.648
LEV	52.431	7.576	0.657	1597.788	0.000	495.578
Rayleigh	19.714		23.669	64.041	0.475	117.357
Weibull	2.151	30.188	18.369	59.095	0.616	69.527
K	0.947	171.624	24.535	34.284	0.999	145.742

Note that, in agreement with the discussion in section 9.3.2.3, the value of N is large for the Rayleigh, Weibull, and K distributions and near 1 for the LEV distribution. Note also, by comparison to figure 44, that the multiple IID copies of lognormal, Rayleigh, K, and Weibull distributions all produce significantly better fits to the upper tail than do the single lognormal or LEV distributions. Hence, the objective for this study was realized.

9.3.2.5 Comparison to Predictions of Physical Model.

Equation 30 represents a distribution of the gated peak-to-peak noise as the maximum of some RVs. In its derivation, Margetan, et al. [42] suggested that the number of discrete RVs should be related to the length of the gate and the duration of the sonic pulse. In particular, if T is the length of the gate and $\Delta\tau$ is a parameter known as the equivalent square wave duration (ESWD), then one expects

$$N = cT/\Delta\tau \quad (33)$$

where c is a numerical proportionality constant and $\Delta\tau$ is defined in the following way. Suppose the pulse has an envelope in the form $A(t)$, which has a peak value A_o . Then

$$\Delta\tau = [\int A^2(t)dt] / A_o^2 \quad (34)$$

For the limited number of cases studied in reference 42, it was found that $c \approx 1$ when one took into account the physically expected variation of RV properties within the time gate.

As a first approximation, consider a Gaussian-shaped pulse, for which it can be shown that

$$\Delta\tau = 0.663 / \Delta f_{6B} \quad (35)$$

where Δf_{6B} is the 6-dB bandwidth. Combining equations 33 and 35 leads to the result

$$N = 1.5c\Delta f_{6B}T \quad (36)$$

Table 9 tests this model by comparing the values of N for the Rayleigh (N_R), Weibull (N_W), and K (N_K) distributions determined for a set of forging data set (which were shown in table 8), to the predictions of equations 35 and 36, based on the nominal values for the probes used, as provided by Howard [56]. One sees that there is reasonable agreement if we take c to have a value on the order of 2. The fact that values of N , which are derived by the maximum likelihood analysis of the noise distribution, correlates well with the values derived from the physical models, provides significant support for the latter. However, further work is required to identify the source of the factor of 2 discrepancy in the value of c observed by Margetan, et al. [42] and in this work.

TABLE 9. COMPARISON OF VALUES OF N COMPUTED FROM NOMINAL PROBE PARAMETERS TO THOSE DETERMINED BY MAXIMUM LIKELIHOOD ANALYSIS OF FORGING NOISE DATA

File	T (μsec)	f _c (MHz)	Δf _{6dB} (MHz)	Δτ (μsec)	T/Δτ	N _R	N _W	N _K
C	8	4.5	2.0	0.33	24	44	35	48
D	4	4.5	2.0	0.33	12	23	18	24
E	2	4.5	2.0	0.33	6	10.8	8.7	11.1
F	1	4.5	2.0	0.33	3	5.4	8.5	6.4
K	8	4.9	2.2	0.30	27	51.6	69	100
L	8	6.6	3.0	0.22	36	63.0	58.9	68.4
M	8	9.2	4.1	0.16	50	-	-	-
N	8	4.5	2.0	0.33	24	33.8	28.3	36.0

9.3.2.6 Conclusions.

The plots and tables presented above are representative exhibits of the analysis results for a billet data set and for a forging data set. The major conclusions based on all data analyzed are:

- Rounding appeared to be the type of interval censoring that had taken place. In any event, the conclusions which follow are reasonably invariant to the actual type of censoring.
- The LEV distribution is *not* a viable candidate for the envelope amplitude distribution. However, it is interesting to note that the estimate of N was near 1 for LEV for all data sets, meaning that the estimate of N for other distributions should be large but not practically infinite. The LEV distribution tends to overestimate the area in the right tail, an observation which has been made before [38].
- The K distribution performed quite well in comparison with the lognormal, Rayleigh, and Weibull; each of which exhibited good performance (results are mixed for billet data with respect to the K, Rayleigh, and Weibull distribution).
- The lognormal distribution produced estimated values of N an order of magnitude smaller than those produced by the Rayleigh, Weibull, and K distributions. The theoretical meaning of this result remains to be discovered.
- The lognormal distribution tends to produce superior results for forging data sets.
- The K distribution does well for billet data sets.
- Data sets that contained only even observations were considered to be invalid and were not used in subsequent analyses.
- GOF (Goodness of Fit) was poor for all distributions when recordings represented mixtures of high- and low-noise regions. In other words, the results of the current analysis should only be applied to noise data taken from a homogeneous region.

9.4 MODELING SIGNAL-PLUS-NOISE DISTRIBUTIONS.

The work done on the implementation in the first generation methodology signal distribution has been enhanced, and this section summarizes the current understanding. At the heart of the POD methodology is the use of physical models to describe the flaw response that would be expected in the absence of noise. The first generation implementation of the methodology (sections 5 through 7) describes the responses from sets of nominally identical FBHs or SHA inclusions. These are used to empirically define the distributions of the model-predicted response, with the emphasis placed on the microstructural contributions to the variability. However, using models to describe the form of this distribution would reduce the need to rely on empirically determined FBH responses.

In section 9.3, the models for distributions of noise were discussed and in section 6, deterministic models for signals in the absence of noise were presented. In this subsection, preliminary results show how these might be combined to predict the microstructural contributions to the signal distribution. These concepts are strongly motivated by approaches used previously in the radar and signal processing communities.

The discussions will be based on some basic assumptions:

1. Signal and noise value add linearly to any transducer position. Therefore, the total signal observed consists of the signal that would occur in the absence of any microstructural-induced noise and the noise waveform that would have existed for that probe position in the absence of a flaw. The latter is not completely true since the volume occupied by the flaw will not generate noise signals. This effect is partially compensated for by the fact that energy reflected from the flaw can backscatter from the microstructure, reflect again from the flaw, and contribute to the backscattered field.
2. In a scanned inspection, the spatial correlation of the noise is on the order of the spatial extent of the flaw response. This implies that, as the beam is scanned over the flaw, there will not be a rapid noise-controlled variation of signal, as shown in figure 49(b), but rather a gradual rise and fall, as shown in figure 49(a). This is consistent with our observations on the FBH and SHA inclusions studied experimentally during the validation studies. Figure 50 presents an example for the latter case. This assumption implies that noise signals will enhance some flaw responses and depress others. It should be noted, that it is inconsistent with the assumption used in Gilmore's suggested noise adjustments of the GE Effective Reflectivity [29] to accommodate the effects of noise, as described in section 8.2.2.2. The assumptions used in that work, that noise always adds to flaw signals, is more consistent with the behavior shown in figure 49(b).

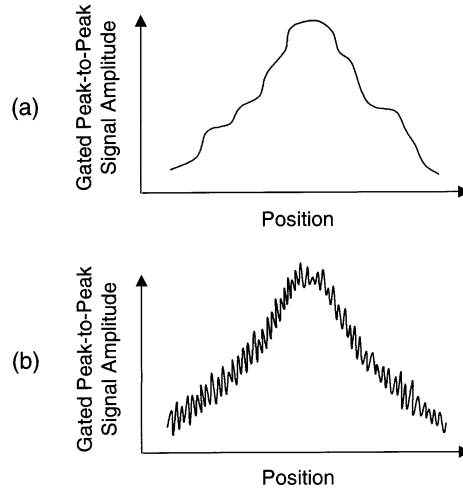


FIGURE 49. SCHEMATIC OF NOISE VARIATIONS IN A LINE SCAN THROUGH FLAW
 (a) CORRELATION LENGTH OF NOISE ON THE ORDER OF FLAW RESPONSE AND (b)
 CORRELATION LENGTH OF NOISE SMALL WITH RESPECT TO FLAW RESPONSE

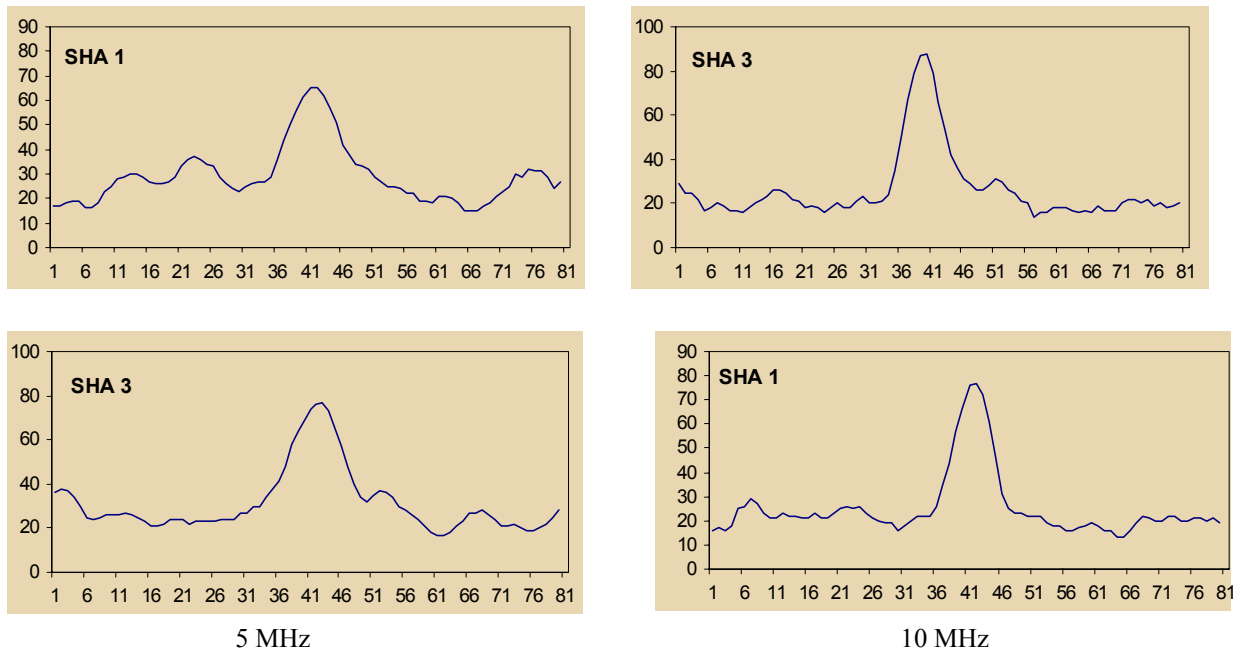


FIGURE 50. EXPERIMENTAL RESULTS OF LINE SCANS THROUGH NO. 2 SHAs
 (Note: SHA n denotes the n th SHA in the total 8 SHAs of identical size. Abscissa is
 in units of 10-mil scan increments.)

A simple physical model can be used to qualitatively indicate the implication of these assumptions. Consider a flaw producing a noise-free response S_o , and assume that it is embedded in an ensemble of noise signals n . Here, S_o and n are considered to be complex

numbers or phasors representing the amplitude and phase of the indicated quantity. Then the total signal, S , is given by

$$S = S_o + n \quad (36)$$

To conceptually understand some of the properties of this addition, one can make the simplifying assumption that all noise signals have constant amplitude, N , but random phases, θ , with respect to S_o ,

$$n = N e^{j\theta} \quad (38)$$

where θ is uniformly distributed between 0 and 2π .

Then simple trigonometry shows that

$$\langle S \rangle = S_o \quad (39)$$

$$\langle S^2 \rangle = S_o^2 + N^2 \quad (40)$$

$$\langle S - S_o \rangle^2 = N^2 \quad (41)$$

These assumptions lead to the idea that the expected magnitude, $\langle S^2 \rangle^{1/2}$, will approach the noise-free signal when the signal-to-noise ratio is large ($S_o/N \gg 1$) and will approach the noise when signal-to-noise ratio is small ($S_o/N \ll 1$). These results seem reasonable and suggest that a more complete model should be examined.

The radar community has studied the case where the signals are narrow-band and detection is based on the envelope in some detail [57]. Suppose that the signal has an envelope A and that the in-phase and quadrature components of the noise are independent and normally distributed with variance σ^2 . Then the density function for the envelope of the signal in the presence of noise is given by the Rician distribution

$$f_{(r)} = \frac{r}{\sigma^2} e^{-\left(r^2 + A^2\right) / 2\sigma^2} I_0\left(\frac{rA}{\sigma^2}\right) \quad (42)$$

where I_0 is the modified Bessel function of the first kind and zero order.

Figure 51 shows the form of this distribution that is observed for various values of the signal-to-noise ratio, $A^2/2\sigma^2$. Examination of equation 42 shows that the Rician distribution approaches the Rayleigh distribution in the absence of signal ($A = 0$), and a normal distribution centered about the noise-free signal for high signal-to-noise ratio [58]. Hence, it is a more rigorous description of the ideas described in equations 39-41.

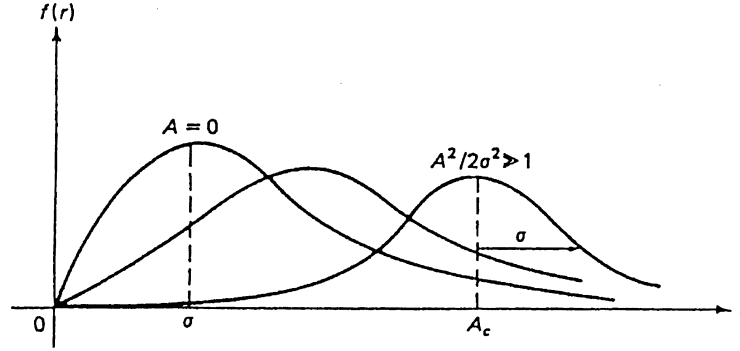


FIGURE 51. RICIAN DISTRIBUTION FOR VARIOUS VALUES OF THE SIGNAL-TO-NOISE RATIO $A^2/2\sigma^2$

Simulations have indicated that this approach has promise for ultrasonic examination of titanium, despite the fact that the assumptions of narrow-band signals and white noise are not rigorously valid [59]. In these simulations, a synthetic flaw signal of variable amplitude was superimposed on 256 noise waveforms were acquired from the same sample of Ti-64 at different positions with a 5-MHz, focused transducer. In each waveform, the flaw was positioned at the same time interval, near the time when a signal would return from the center of the focal zone and, hence, near the time of peak root-mean-square noise. The peak-to-peak signal in an 0.75- μ sec gate was then computed. The distribution of peak-to-peak signals was compared to the predictions of the Rician Theory, as described further below.

Figure 52 shows the noise-free value of the signal and its form when a significant amount of noise is added. In this example, the signal and noise are in phase so that the signal is enhanced. However, a suppression of the signal is just as likely. Because of the finite length of the gate and the fact that the signal is broadband, it would not be obvious what the appropriate values of A and σ would be if they were inserted into equation 42. In the initial analysis, A was chosen as the average of the magnitudes of the two largest cycles in the synthetic flaw signal with σ^2 being taken as the rms value of the noise data.

When the predictions of equation 42, with the parameters evaluated in this way, were compared to the simulated data, good fits were obtained for high signal-to-noise ratios but not for low signal-to-noise ratios. This difficulty was related to the finite length of the gate. As discussed by Margetan, et al. [42] and in section 9.3, the distribution of gated peak-to-peak noise signals is best described by independently sampling the envelope distribution several times and selecting the maximum value. The physical rationale is that there are multiple independent opportunities, N , to sample the noise in the gate, which are related to the ratio of gate length to pulse length as described in equation 33. The same ideas should apply to the distribution of gated peak-to-peak flaw signal. However, in this case, the noise distribution would be sampled in some cases and the flaw distribution in other cases. To see if the Rician distribution was applicable in the time interval in which the flaw signal was present, the gate length was reduced to the duration of one independent event, 0.19 μ sec in this case. The simulation results and the Rician predictions were then found to be in good agreement.

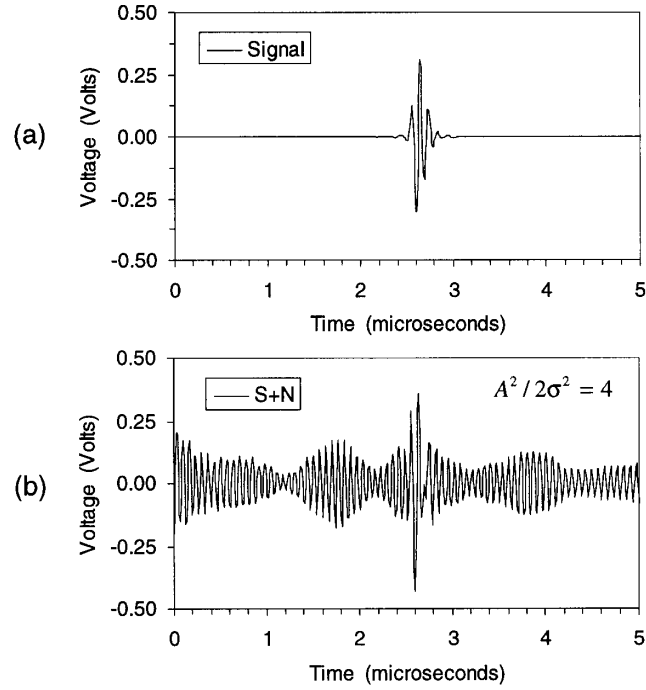
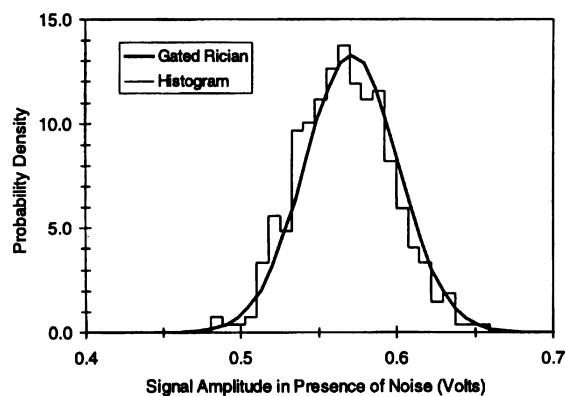


FIGURE 52. SYNTHETIC FLAW SIGNAL (a) NOISE FREE AND (b) $A^2/2\sigma^2 = 4$

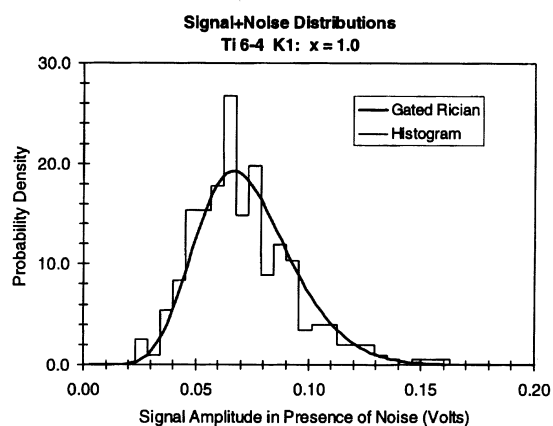
Based on these results, the model for gated peak-to-peak noise [42] was used for the case where a flaw signal was present. The result of this extension is the model for signal plus noise distributions that is the subject of this section. The essential idea was to again view the detection process as the independent sampling of N random variables. N minus 1 ($N-1$) of these were taken to be the noise in the absence of a flaw, and are described by the Rayleigh distribution. In the last one, equation 42 is used to describe the addition of signal and noise. N is given by equation 33.

Figures 53 and 54 present, respectively, a comparison of the predictions of this model to data simulated as previously described (figure 53) and the experimental data for one of the SHA samples (figure 54). Figure 53 compares the results of this simulation (abruptly changing curve labeled histogram) to the gated Rician model for signal plus noise distribution (smooth curve) where the parameters have been selected as discussed above, where c is the constant first appearing in equation 33, which denotes N (taking $c = 1$). X is a measure of the noise-to-signal ratio¹⁹. It can be seen that the model smoothly fits the histogram data reasonably well for all the signal-to-noise ratios examined. In the comparison to experimental data in figure 54, the solid line represents the predicted distribution. There are only eight observations where the responses are nominally identical to no. 3 SHAs, so a meaningful experimental histogram cannot be constructed. However, to qualitatively indicate that the data are consistent with the theory, square symbols have been placed on the abscissa at the discrete values of the responses from the eight nominally identical flaws. Statistical tests of the goodness of fit have not yet been performed, but the fact that the discrete points generally fall within the major portion of the distribution is taken as qualitative, experimental confirmation of the theory.

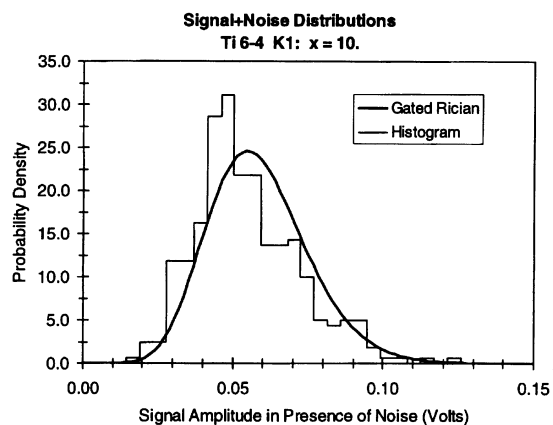
¹⁹ Low values imply high SNR and high values imply low SNR.



$x=0.1$ (good SNR)



$x=1.0$ (SNR~1)



$x=10$ (very low SNR – essentially noise)

FIGURE 53. A COMPARISON OF SIMULATED DISTRIBUTIONS OF GATED PEAK-TO-PEAK, SIGNAL-PLUS-NOISE DISTRIBUTIONS TO PHYSICALLY BASED MODEL PREDICTIONS IN Ti 6-4, SAMPLE K1

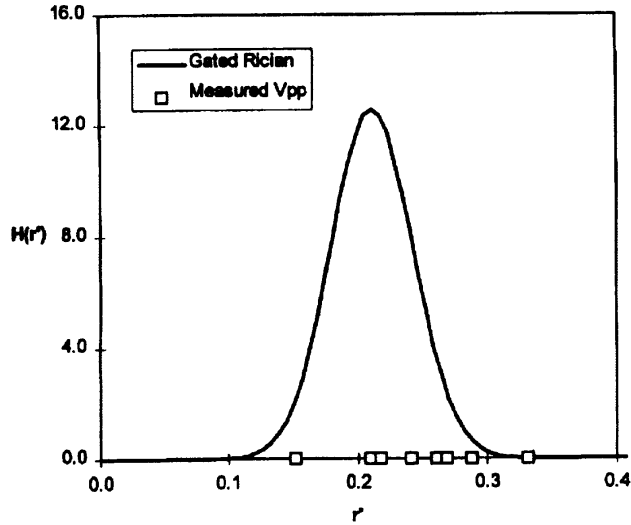


FIGURE 54. A COMPARISON OF THE DISCRETE RESPONSES OF EIGHT SHAs TO THE PHYSICALLY BASED MODEL PREDICTIONS OF THE DISTRIBUTION

It should be emphasized that when the distribution is obtained in this fashion it is difficult to determine whether the threshold is exceeded by the signal from a flaw or a microstructural inhomogeneities. Hence, integrating this portion of the distribution above a fixed threshold would yield POI rather than POTD (see the definition of terms in subsection 5.2.12).

9.5 BEAM MODULATION EFFECTS.

The discussion in section 9.4 presumes that the only source of microstructure-induced variation between the signals from nominally identical flaws is additive random noise. Another possible source is the fact that phase fluctuations can develop in an ultrasonic beam as it propagates through titanium [3, 59, and 60], which can lead to significant fluctuations in back surface signals and, presumably, flaw signals. This is a subject that is still under study in the Fundamental Studies Task. At the present time, it is handled empirically in the methodology. A future goal should be to develop physical models that take this phenomenon into account.

9.6 EFFECTS OF SURFACE ROUGHNESS.

Surface roughness has the potential to limit the detection of hard-alpha defects in titanium billets. Qualitatively, surface roughness can deviate, attenuate, and fragment the incident beam pattern, which results in the degradation of the SNR. These effects are somewhat analogous to the beam modulation effects discussed above. Prior to this program, limited theory was available to quantitatively estimate the effects of surface roughness on the SNR. Accordingly, a general theory was developed to determine the effects of surface roughness on (1) signals generated by subsurface (hard-alpha) inclusions, (2) the noise generated by the microstructure of the material, and (3) the ratio of the signal to the noise. The effects of surface roughness on hard-alpha detection were estimated from this general formulation. This study did not consider additional noise that might be generated by the billet surface.

9.6.1 Measurements of Surface Roughness.

To quantify the surface roughness on representative samples of billets, topographic measurements were made at the General Electric Corporate Research and Development (GECRD) laboratories. A small region of the surface of billets was obtained from two manufacturers. For two samples, the root-mean-square deviation of the roughness was approximately 35 μm , which is sufficiently close to the wavelength of sound in water: $\lambda=300$ μm at 5 MHz and 150 μm at 10 MHz to be of concern. Roughness of this magnitude has the potential to have significant effects on the shape of the ultrasonic beam propagating into the billet. The theories described below were used to quantify this effect.

9.6.2 Theory for the Effects of Surface Roughness on Grain Noise.

The first theory that was considered was the effects of surface roughness on backscattering from the microstructure. A multitude of behaviors were found that heavily depended upon the rms height and correlation length of the surface roughness as well as the average size of the grains. Formulas were developed that showed how the change in the normalized rms noise (the noise in the presence of roughness divided by the noise in the absence of roughness) depended on these parameters [61-64].

Refer to reference 61 to obtain a fundamental model for the interaction of the rough surface with grain noise (an analytic series solution for the roughness-induced change in the grain noise was developed). It was found that surface roughness can either enhance or decrease the grain noise, depending on the size and focal length of the transducer, the frequency, and the depth beneath the surface. Experimental measurements conducted by Peter Nagy at Ohio State University were compared to the predictions of the new theory and found to be in good agreement. To test and validate this series solution, a numerical and more exact solution to the same problem was developed [62]. The analytic series solution was also improved by allowing the microstructure to have a finite correlation length [64]. Using these models, it was discovered that normalized grain noise can be reduced (sometimes greatly) by using focused probes and measuring the noise at the focal depth of the probe [63]. This important result has striking consequences on the normalized SNR—and leads to a recommendation by the authors that the SNR be measured “on the fly.” By this, it is meant that the noise should be determined from acoustic data obtained in the vicinity of the flaw. More detail will be given in section 9.6.4.

9.6.3 Theory for the Effects of Surface Roughness on the Signal From Subsurface Inclusions.

A hard-alpha inclusion presents a weak ultrasonic contrast with the titanium matrix. Since the contrast could be diminished further by surface roughness, the effects of surface roughness on the signal for a subsurface inclusion was calculated in the weak scattering limit [65-71]. Surface roughness was found to be particularly deleterious when the scatterers are near the surface. In addition, it was shown how rough surfaces affected the signal from unvoided and uncracked worst-case hard-alpha inclusions.

References 65, 66, and 69 discuss the effects of surface roughness on the signal, which comes from a subsurface inclusion. An important factor is that the ultrasonic wave must pass through

the surface once on its way from the transducer to the flaw and then again on its way back to the transducer [65, 66, and 67]. This double transmission introduces important correlations in the phase and leads to a near-surface dead zone where the signal is greatly attenuated. Experimentally, it was also found and demonstrated that double reflection can be used to determine the correlation length and rms height of the rough surface.

Roughness changes the rms signal from a subsurface flaw and introduces a random component in the signal. To study this effect, a series solution was introduced for the scattering from a ellipsoidal weak scatterer beneath a rough surface with a Gaussian profile in reflectivity (representing the diffusion zone). This series solution showed how the average and the variance of the signal depended the rms height and correlation length of the surface roughness as well the frequency and the radius of the transducer and its focal length [68 and 71]. It was also shown that the a frequency-dependent transmission constant is useful on a very wide range of experimental conditions, including those that occur in hard-alpha inspections in describing the effects of surface roughness on the average signal [70].

The possible existence of unvoided and uncracked hard-alpha inclusions is one of the nightmares of hard-alpha inspection. Such an inclusion might consist of a central hard-alpha nugget and a more or less extended diffusion zone. The ultrasonic reflection from such an inclusion depends crucially on the exact velocity profile of the inclusion. There are two limiting cases. The best case occurs when the velocity profile changes abruptly at an internal interface between the hard-alpha nugget and the extended diffusion zone. This abrupt change in properties leads to a relatively strong echo that increases with frequency. The worst-case scenario is that the velocity changes continuously and smoothly from the center of the hard-alpha nugget through the diffusion zone and then through the host metal. In this case, there is no strong echo. Furthermore, as the frequency is increased, the incident wave is guided through the inclusion with little back scattering; i.e., a large worst-case inclusion would be essentially invisible at 5 MHz. Additionally, surface roughness can further degrade the detectability of such worst-case inclusions. Table 10 presents a guide showing just how bad the situation would be if hard-alpha inclusions were of this type. The left three columns of table 10 consider the case of a smooth surface and the right three consider the case of a rough surface. Here h is the rms height of the part surface roughness, B , is the e^{-1} radius of the inclusion (which has assumed to have a Gaussian variation in velocity to represent a diffusion zone), and f^{opt} is the frequency at which the detectability is optimum [68].

TABLE 10. OPTIMUM FREQUENCY FOR DETECTION OF WORST-CASE HARD-ALPHA INCLUSIONS BENEATH SMOOTH (FIRST THREE COLUMNS) AND ROUGH (SECOND THREE COLUMNS) SURFACES

$h(\mu\text{m})$	$b(\text{mm})$	$f^{opt}(\text{MHz})$	$h(\mu\text{m})$	$b(\text{mm})$	$f^{opt}(\text{MHz})$
0.0	5.0	0.14	20.0	5.0	0.14
0.0	1.0	0.71	20.0	1.0	0.71
0.0	0.5	1.42	20.0	0.5	1.41
0.0	0.1	7.09	20.0	0.1	5.97

From these results for the worst-case hard-alpha inclusion, it can be seen that there is a tradeoff between the ultrasonic inspection frequency and the size of the hard-alpha inclusions for the maximum detectability of the inclusion. Most importantly, the detectability of an uncracked hard alpha that has a diffusion zone represented by the Gaussian variation in velocity appears to decrease for large flaws. The presence of rough surfaces degrades the situation by attenuating the signal and inducing a near-surface dead zone.

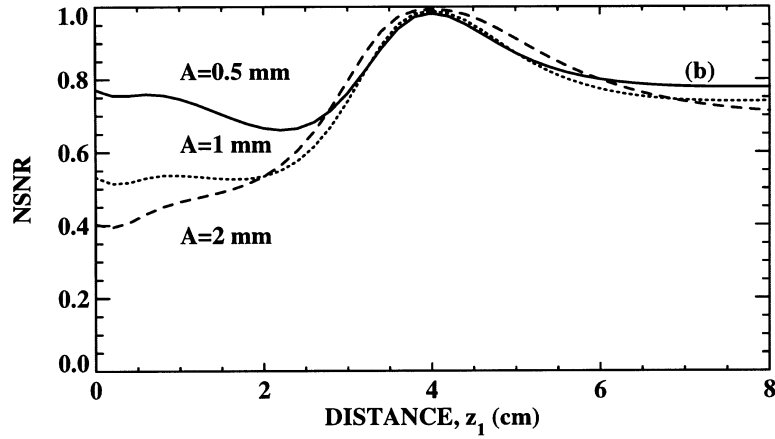
Special care is suggested in selecting the frequency for the inspection of parts with rough surfaces. Best results are obtained at the optimal frequency, f^{opt} , which is inversely proportional to the size of the inclusion. At low frequencies, the detection of relatively large inclusions is maximum; small inclusions are obscured, and the surface roughness is less important. At high frequencies, large inclusions become invisible, but the detectability of smaller inclusions increases. However, arbitrarily large frequencies cannot be used because surface roughness losses and grain scattering increase with increased frequency.

In the conduct of this study, it was recognized that most hard-alpha inclusions are voided, a condition that leads to quite different conclusions. Moreover, if the velocity does not vary smoothly, e.g., as in the “best case” described above, the results would also be different. Analysis of the data obtained in the Contaminated Billet Study (described in section 10.2), will be an important step towards determining the possibility that such a worst case inclusion can occur.

The specific case of the detectability of inclusions in titanium billets was addressed in detailed numerical calculations that started from the topographic measurements (see section 9.6.1) of the rough surface performed at GECRD [72]. The transducer was modeled using the parameters for the multi-zone lens system in use at GE [73]. The billet was modeled as an infinitely long cylinder whose surface had the roughness of the actual billets measured by GE. The fields inside the billet were then calculated using codes that had been previously developed for multizone lens design [74]. The roughness induced change in the signal strength was estimated by observing the field strength calculated in the focal zone. A predicted change of several dBs was found in the signal for the roughest surface measured by GE. These results are consistent with estimates obtained from the series expansion solutions mentioned above.

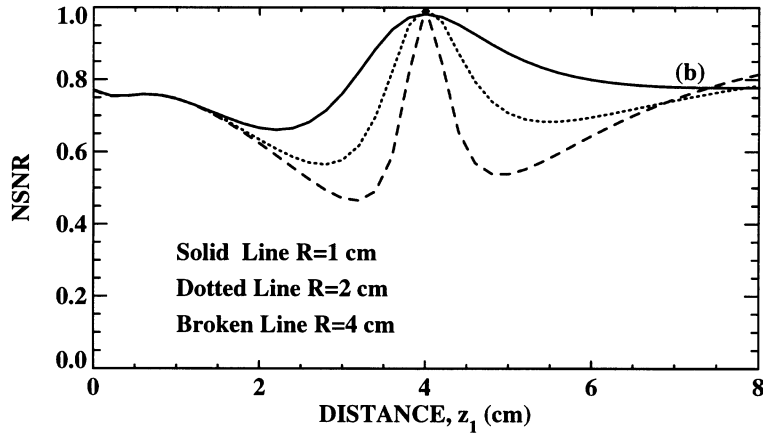
9.6.4 Predictions of the Effects of Surface Roughness on the SNR.

The SNR for an inclusion beneath a random rough surface was analyzed and predicted, including the dependence of the normalized SNR on the rms height and correlation length of the surface; on the size of the microstructure; and on the frequency, aperture, and focal length of the transducer [75]. This normalized SNR is defined as the ratio of the SNR, taking the surface roughness into account, to the SNR neglecting that roughness. Figure 55 shows, for a particular surface condition, how the normalized SNR depends on the size of the scatterer, while figure 56 shows its dependence on the size of the transducer. The most important result is that focused probes reduce the effects of surface roughness on the SNR if the noise is measured on the fly, i.e., the noise should be determined from the acoustic sample that contains the flaw signal. Simple engineering formulas have been developed that allow the reader to estimate the signal-to-noise ratio [75].



(Note: The surface correlation length is $l_s = 1.5$ mm, the rms height is $h = 0.02$ mm, the correlation length of the microstructure is $l_m = 0.1$ mm, while the inspection frequency is $f = 10$ MHz, and the radius of the transducer is $r = 10$ mm.)

FIGURE 55. NORMALIZED SNR (NSNR) AS A FUNCTION OF DEPTH FOR DIFFERENT RADII INCLUSIONS ($A = 0.5$, 1.0 , AND 2.0 mm) FOR A FOCUSED TRANSDUCER



(Note: The radius of the inclusion is $a = 0.5$ mm, the surface correlation length is $l_s = 1.5$ mm, the rms height is $h = 0.02$ mm, the correlation length of the microstructure is $l_m = 0.1$ mm, and the inspection frequency is $f = 10$ MHz.)

FIGURE 56. NORMALIZED SNR AS A FUNCTION OF DEPTH FOR DIFFERENT TRANSDUCER RADII ($R=10$, 20 , AND 40 mm) FOR A PROBE FOCUSED AT 40 mm INSIDE THE METAL

This formalism was used to estimate the effects of surface roughness on SNR for hard-alpha inclusions [76]. The rms height was chosen to be 35 μm with correlation lengths of 1.25 cm and 3.75 cm—to approximately agree with the topography measurements made at GECRD. It was found that both the signal and the noise were attenuated by several dBs. However, the normalized SNR was predicted to be unchanged if the noise was measured on the fly.

9.6.5 Summary of Crucial Results.

Computable theories were developed to determine the effects of surface roughness on the signal from subsurface inclusions, the noise generated by the material's microstructure, and the SNR. All of these theories were developed and put into a computable form. The crucial variables that enter into the formalism were identified as the rms height and correlation length of the surface roughness, the frequency and focal length of the transducer, as well as the size and position of the scatterer.

These models were used to estimate the importance of surface roughness. One important figure of merit was found to be the decrease in the average intensity of the beam. At 5 MHz, the average beam intensity decreased by 4 dB for a 35- μm rms roughness. Due to the greater ratio of surface roughness to wavelength, the decrease was 15 dB at 10 MHz. More sophisticated calculations give similar (although somewhat smaller) estimates for the decrease in the average signal strength. Thus, at 5 MHz the attenuation of the average signal is the same size as other variations in the theoretical model. On the other hand, if a new system is developed and operated at 10 MHz, then surface roughness may well be the limiting factor in the detection of hard alpha when billets are fabricated.

One of the major results of this study was the discovery that the deleterious effects of surface roughness on the SNR can be significantly reduced for focused probes if the noise is measured on the fly. Model studies were conducted for the signal, the noise, and the signal-to-noise ratio for simulated hard-alpha inclusions in 5" diameter billets with rough surfaces similar to those reported by GECRD. A sophisticated model predicted that both the signal and the noise are attenuated by a significant amount, 2-3 dB for 35- μm rms surface roughness. However, the model also showed that the signal-to-noise ratio was nearly unaffected if the noise and signal were estimated from the same acoustic sample, i.e., the noise is measured on the fly. These results have important consequences for the detection of hard alpha using the multizone system. GECRD strongly suggests that the signal-to-noise ratio should be calculated on the fly and that this local signal-to-noise ratio should be used as the most important indicator of the presence of a flaw. Other indicators may not work as well. For example, suppose a threshold value is set for the signal to detect hard alpha. Since surface roughness attenuates the average signal, certain flaws would not be detected by a fixed threshold but would be detected by an on-the-fly measurement of the signal-to-noise ratio. The same objection holds if the noise is measured in only a few positions and a global signal-to-noise ratio is used for detection. However, this objection is overcome by measuring the signal-to-noise ratio on the fly, since both the signal and the local noise are attenuated in the same way. Based on these model predictions, it is clear that detectability is controlled by such a local measure of signal-to-noise ratio. It will be a challenge

for the future to incorporate this insight into improved detection algorithms and the assessment of their capability.

9.6.6 Conclusions.

The rough surfaces characterized by GECD change the predicted signal, at most, by 2 or 3 dBs for inspections such as those made with the multizone system operated at 5 MHz. Even these modest effects can be removed if the SNR is measured on the fly. Hard-alpha detection will be degraded for rougher surfaces (or for measurements made at higher frequencies). In these cases, it is recommended that noise be measured on the fly.

9.7 STRATEGIES FOR INCORPORATING PHYSICS-BASED DESCRIPTIONS OF NOISE AND SIGNAL-PLUS-NOISE IN THE NEW METHODOLOGY.

In the initial implementation of the new methodology, the noise distribution was obtained empirically based on an assumed lognormal distribution. The flaw response was determined from a physical model that would be exhibited in the absence of microstructure. A statistical model was developed to describe the microstructure-induced deviations from this response based on the response measurements of a number of nominally identical reflectors.

The physics-based models for the form of these distributions can be used in future versions of the methodology in several ways. The noise models presented in subsection 9.3 provide a better description of the large amplitude tail of the distribution than the currently used lognormal distribution. Where issues such as PFA or SNR are of concern, these physics-based models should improve results when fitted to experimental data.

In some cases, one would like to estimate the POD and PFA without obtaining extensive empirical data. Figure 57 schematically illustrates how this could be done. Based on a few microstructural parameters such as the figure of merit (FOM) for noise generation, the microstructural contributions to the noise and signal distributions could be determined based on the approaches described in sections 9.3 and 9.4. By incorporating the dependence of the flaw response on instrumentation parameters, the broadening of the signal distribution due to scanning could be added. Flaw morphology effects would be added based on future work, as noted in section 10.

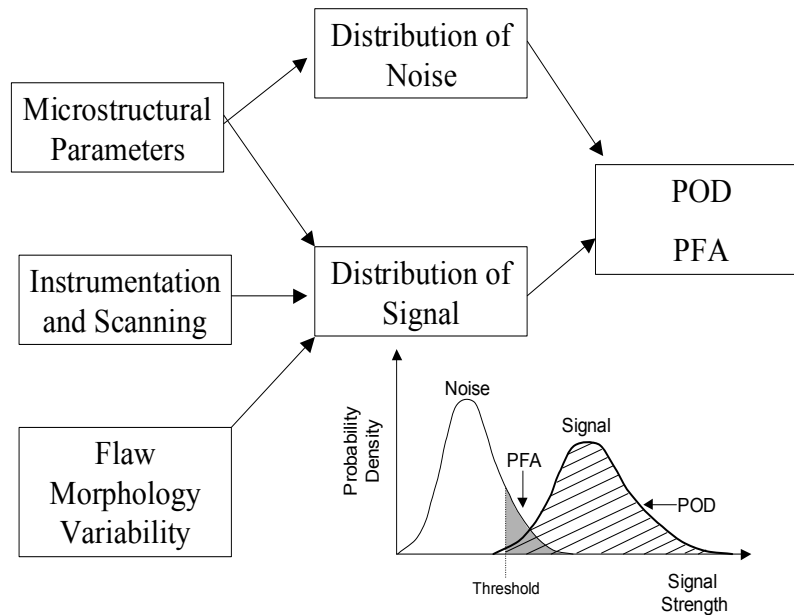


FIGURE 57. STRATEGY OF INCORPORATING PHYSICS-BASED DESCRIPTIONS OF SIGNAL AND NOISE DISTRIBUTIONS IN POD METHODOLOGY

10. FUTURE DIRECTIONS.

10.1 RANDOM DEFECT BLOCK.

At the conclusion of this work an important experiment was in progress, which was designed to provide a basis for testing the validity of the methodology. Diffusion-bonding techniques had been used to construct a section of a billet on the order of 4 feet in length that contained a large number (on the order of 50) SHAs. These consisted of right circular cylinders, with axes roughly parallel to the axis of the billet, and spheres. The sizes, nitrogen contents, and inclinations had been selected to simulate a wide range of detection opportunities, presenting signal-to-noise ratios expected to cover a broad range [77]. Within certain constraints, the SHAs positions were random within the billet.

This billet was to be scanned using both conventional and zoned techniques, leading to a set of response and signal-to-noise ratio measurements for each defect detected. These observations were to be compared to the predictions of the physical model. These experiments, conducted after the completion of the work but before the finalization of this report, demonstrated that the methodology requires further refinement to accurately predict the detection probabilities of defects in a realistic industrial environment. In particular, they showed the need to “productionize” the methodology to account for a number of input parameters in the physics-based models that are not fully controlled in an industrial environment and, thus, represent a source of additional variability. For example, there will be a range of transducer characteristics that are consistent with calibration tolerances and the noise level will vary from billet to billet and from region to region within a billet. Consequently, there will be additional variability associated with setup procedures, e.g., the inspector’s ability to reproducibly align the probes.

Follow-on work is planned to develop approaches to introduce these as well as any other sources of variability into a POD prediction that is fully representative of the industrial setting.

10.2 CONTAMINATED BILLET STUDY.

Because of a power failure during the processing of an ingot of nonrotor grade material, a set of billets has become available that contains over 60 indications that appear to be associated with hard-alpha inclusions. These are being used in a variety of tasks ranging from studies of the deformation of hard-alpha inclusions during forging to improving the understanding of their detectability [78]. In the latter context, ten defects have been set aside for a detailed analysis, which include carefully documenting their ultrasonic response, determining the morphology of any pores and the spatial extent of enhanced nitrogen regions, extracting from this information models of the spatial variation of elastic constants and density, and using these to drive models predicting the response of the naturally occurring hard-alpha inclusions. Good agreement will represent further validation of models. Preliminary descriptions of the models and their validation can be found in reference 79. This data will also significantly enhance the understanding of the relationship of flaw morphology to signal strength. This information that will become an essential ingredient in estimating the detectability of naturally occurring hard-alpha inclusions.

10.3 PORTABLE POD.

One of the major secondary objectives in the development of the new methodology was the development of a capability to provide estimates of the effects of changes in inspection procedures without requiring a new set of samples and experiments each time, i.e., the development of a so-called “portable” POD. The basic ingredients of this idea are contained in section 5.2.8. This shows that the POD for a given set of experimental conditions can be determined by an integral over a basic POD, which contains variables associated with those conditions as parameters. The further development of this idea will be an important topic of future work.

10.4 ADJUSTING FOR NATURAL-FLAW PARAMETERS.

The POTD predictions reported in this document are for FBHs and SHAs. As indicated in figures 9 and 57, additional work is required to introduce the effect of the morphology of naturally occurring defects on the signal distribution. The Contaminated Billet Study will be an important source of data for this purpose. The definition of the detailed strategies is currently in progress.

11. SUMMARY AND CONCLUSIONS.

11.1 SUMMARY.

This report has described a new methodology for determining POD. Based on statistical detection theory, the underlying strategy was to determine distributions of signal and noise, from which the POD, PFA, and ROC curves can be determined. Heavy use was made of physics-based models during the inspection process to minimize the amount of empirical data that must

be gathered. Although motivated by and applied to the case of the ultrasonic detection of internal inclusions in aircraft engine rotating components, the general approach is applicable to a much broader set of cases. This report includes a general review of various methodologies for determining POD, a detailed discussion of the new approach, results of its application to the ultrasonic detection of flat-bottom holes and synthetic hard-alpha inclusions in flat plates under laboratory conditions, and a comparison of these results to those obtained by other methodologies. Ongoing experiments and analysis, aimed at validation of the methodology and at extending its predictions to the detection of naturally occurring flaws, are also summarized. These experiments, completed after the work described herein but before the finalization of this report, show that the methodology needs to be “productionized” to account for a number of input parameters in physics-based models are not fully controlled in an industrial environment and, thus, represent a source of additional variability.

As currently implemented, the methodology used physical models to describe the response of flaws in the absence of microstructural effects and the variations in this response produced by scan parameters. By conducting limited empirical experiments in which the response of a number of identical targets are measured, the effects of microstructure and other sources of variability are inferred. New tools that will further reduce the need for empirical experiments, based on physical models for the effects of microstructure on the ultrasonic responses, are also included. The output of these tools is a distribution of possible signals rather than a deterministic signal strength.

11.2 CONCLUSIONS.

The new methodology has generated results that are in good agreement and verifies those obtained with existent methodologies for the case of flat-bottom holes in flat plates under laboratory conditions. When applied to synthetic hard-alpha inclusions, it also provides results that are credible. This illustrates the robustness of the technique, since there were difficulties with applying existing methodologies to the same synthetic hard-alpha inclusion data. An experiment is currently in progress to validate these predictions.

The primary initial motivation for the work was to develop an approach to make a probability of detection determination for internal flaws under conditions where other approaches were considered unsatisfactory. However, in the course of this work, a number of other advantages of the methodology have been observed. Because physical models are used during the inspection process, fewer samples are needed than in purely empirical methods, such as the \hat{a} versus a approach, which have significant economic implications. It has not been fully established how sample requirements of the new approach compare to earlier model-based approaches, such as the effective reflectivity (R_e) method. An explicit output of the new approach is a prediction of the probability of false alarm as well as the probability of detection which allows relative operating characteristic curves to be determined. This is done so that the tradeoff between probability of detection (safety) and probability of false alarm (cost) of various candidate thresholds can be evaluated.

Although not explicitly demonstrated in this program, it is anticipated that the new methodology will have other important benefits. By using physical models during the measurement process, it should be possible to:

- estimate the probability of detection (POD) in physical situations where samples are not available or in candidate inspections that have not yet been implemented in hardware, producing a “portable” POD.
- estimate the influence of part geometry on the POD.
- generate information that will allow a more accurate analysis of the effects of various service and inspection scenarios on part lifetimes.

There may also be important implications during service, when new inspections must be developed in response to unanticipated damage mechanisms and POD estimates are needed in a very short time frame. In all of these cases, there are significant economic implications, since there would be a reduced need to make a new set of POD samples every time a new part geometry, material, or failure mode is encountered in applications.

Ultimately, the community must develop a better understanding of the POD of naturally occurring flaws as well as the POD of flat-bottom holes or synthetic hard-alpha inclusions. The availability of validated models that include the effects of measurement geometry on the flaw response will be a significant aid in achieving this objective, since it will allow the responses of flaws from different regions of a part to be compared on an equal footing, thereby allowing all data to be aggregated and analyzed in a uniform way.

12. REFERENCES.

1. J.G. Costa, R.E. Gonzalez, R.E. Guyotte, D.P. Salvano, T. Swift, and R.J. Koenig (Chair), “Titanium Rotating Components Review Team Report,” Federal Aviation Administration, Aircraft Certification Service, Engine and Propeller Directorate, Boston, MA, December 14, 1990.
2. M.F.X. Gigliotti, R.S. Gilmore, and L.C. Perocchi, “Microstructure and Sound Velocity of Ti-N-O Synthetic Inclusions in Ti-6Al-4V,” *Metallurgical and Materials Transactions A*, Vol. 25A, pp. 2321-2329, 1994.
3. F.J. Margetan, P.D. Panetta, and R.B. Thompson, “Ultrasonic Signal Attenuation in Engine Titanium Alloys,” in *Review of Progress in Quantitative NDE*, Vol. 17B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1469-1476, 1998.
4. A.P. Berens, “NDE Reliability Data Analysis,” in *Metals Handbook, Ninth Edition, Vol. 17, Nondestructive Evaluation and Quality Control*, ASM, Metals Park, OH, 1989, pp. 689-701.

5. R.C. Schwant and D.P. Timo, "Inspection and Life Evaluation of General Electric Turbine Rotors," ASME paper 84-JPGC-Pwr-4, 1984.
6. W.D. Rummel, "Recommended Practice for a Demonstration of Nondestructive Evaluation (NDE) Reliability on Aircraft Production Parts," *Materials Evaluation*, Vol. 40, pp. 922-932, August 1982.
7. A.P. Berens, et al., "User's Manual for Probability of Detection Software System (POD/SS)," University of Dayton Research Institute Report UDR-TR-88-12, 1988.
8. D.J. Sturges, R.S. Gilmore, and P.W. Hovey, "Estimating Probability of Detection for Subsurface Ultrasonic Inspection," in *Review of Progress in Quantitative NDE*, Vol. 5A, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 929-936, 1986.
9. J. & H. Krautkrämer, "Ultrasonic Testing of Materials," Springer-Verlag, 4th Edition, 1990.
10. D.O. North, "Analysis of Factors Which Determine Signal-to-Noise Discrimination in Radar," Report PTR-6c, RCA Laboratories, 1943.
11. R.J. Urick, "Principles of Underwater Sound," McGraw-Hill, 1983.
12. W.W. Peterson and T.G. Birdsall, "The Theory of Signal Detectability," Technical Report No. 13, Electronic Defense Group, University of Michigan, 1953.
13. C. Nockermann, G.-R. Tillack, and H. Wessel, "Reliability in NDT: ROC Study of Radiographic Weld Inspections," *NDT&E International*, Vol. 24, pp. 235-245, 1991.
14. J.N. Gray, T.A. Gray, N. Nakagawa, and R.B. Thompson, "Models for Predicting NDE Reliability," *Metals Handbook, Ninth Edition*, Vol. 17, pp. 702-715.
15. F.W. Spencer, "Identifying Sources of Variation for Reliability Analysis of Field Inspections," RTO Workshop 2 – Airframe Inspection Reliability Under Field/Depot Conditions, May 1998, Brussels, Belgium.
16. F.W. Spencer, "Fitting Probability of Detection Curves to Single Inspection Hit/Miss Data," presented at 1998 ATA Nondestructive Testing Forum, Albuquerque, NM.
17. C.-P. Chiou, F.J. Margetan, and R.B. Thompson, "Ultrasonic Signal Characterizations of Flat-Bottom Holes in Titanium Alloys: Experiment and Theory," in *Review of Progress in Quantitative NDE*, Vol. 14B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 2121-2128, 1995.
18. C.-P. Chiou, F.J. Margetan, and R.B. Thompson, "Development of Ultrasonic Models for Hard-Alpha Inclusions in Titanium Alloys," in *Review of Progress in Quantitative NDE*,

- Vol. 16B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1529-1536, 1997.
19. W.Q. Meeker, R.B. Thompson, C.-P. Chiou, S.-L. Jeng, and W.T. Tucker, "Methodology for Estimating Nondestructive Evaluation Capability," in *Review of Progress in Quantitative NDE*, Vol. 15B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1983-1991, 1996.
 20. C.-P. Chiou, F.J. Margetan, and R.B. Thompson, "Modeling of Ultrasonic Signals From Weak Inclusions," in *Review of Progress in Quantitative NDE*, Vol. 15A, D.O. Thompson and D. E. Chimenti, eds., Plenum Press, NY, pp. 49-55, 1996.
 21. W.Q. Meeker, S.-L. Jeng, C.-P. Chiou, and R. B. Thompson, "Improved Methodology for Predicting POD of Detecting Synthetic Hard Alpha Inclusions in Titanium," in *Review of Progress in Quantitative NDE*, Vol. 16B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 2021-2028, 1997.
 22. P. Sarkar, W.Q. Meeker, R.B. Thompson, T.A. Gray, and W. Junker, "Probability of Detection Modeling for Nondestructive Testing," in *Review of Progress in Quantitative NDE*, Vol. 17B, D.O. Thompson and D. E. Chimenti, eds., Plenum Press, NY, pp. 2045-2052, 1998.
 23. W.J. Dixon & F.J. Massey Jr., "Introduction to Statistical Analysis," McGraw-Hill, 1969.
 24. "Non-Destructive Evaluation System Reliability Assessment," Military Handbook 1823.
 25. A.P. Berens, "NDE Reliability of the ECII Series E Eddy Current Inspection System," Report UDR-TR-95-45, 1995.
 26. See Reference 24, section IV.6.4.
 27. See Reference 24, Appendix G.3.2.1.
 28. See Reference 24, Appendix H.3.1.1.
 29. R.S. Gilmore, Corporate Research and Development, General Electric Corporation, private communication.
 30. C.-P. Chiou, Center for Nondestructive Evaluation, Iowa State University, private communication, 1996.
 31. R.H. Burkel, D.J. Sturges, W.T. Tucker, and R.S. Gilmore, "Probability of Detection for Applied Ultrasonic Inspections," in *Review of Progress in Quantitative NDE*, Vol. 15B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1991-1992, 1996.

32. W.T. Tucker, "Likelihood Ratio Methods for Ti Inspections," *Proceedings of the First FAA Inspection Reliability Workshop*, 1993.
33. P.J. Howard and R.S. Gilmore, "Flaw Detection Using Dynamic Thresholding in C-Scan Images with Variable Noise," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 13B, D.O. Thompson and D. E. Chimenti, eds., Plenum Press, NY, pp. 1745-1752, 1994.
34. D.C. Copley and P.J. Howard, "Operational Performance of a Multizone Billet Inspection System," in *Review of Progress in Quantitative NDE*, Vol. 14B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 2145-2151, 1995.
35. W.L. Tucker, General Electric Internal Report (unpublished), 1992.
36. P.J. Howard, R.H. Burkel, and R.S. Gilmore, "The Statistical Distribution of Grain Noise in Ultrasonic Images," in *Review of Progress in Quantitative NDE*, Vol. 15B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1517-1524, 1996.
37. P.H. Wirsching and J.R. Carlson, "Model Identification for Engineering Variables," *Journal of Engineering Mechanics Division, ASCE*, Vol. 103, No. EM1, February, 1977.
38. P.J. Howard, R.H. Burkel, and R.S. Gilmore, "The Statistical Distribution of Grain Noise in Ultrasonic Images," in *Review of Progress in Quantitative NDE*, Vol. 15B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1517-1524, 1996.
39. M.R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York, 1983.
40. J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, John Wiley & Sons, New York, 1978.
41. W.T. Tucker, "Likelihood Ratio Methods for Ti Inspection," *Proceedings of the First FAA Inspection Reliability Workshop*, 1993.
42. F.J. Margetan, I. Yalda, and R.B. Thompson, "Predicting Gated-Peak Grain Noise Distributions for Ultrasonic Inspections of Metals," in *Review of Progress in Quantitative NDE*, Vol. 15B, D. O. Thompson and D. E. Chimenti, eds., Plenum Press, NY, pp. 1509-1516, 1996.
43. H.L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley & Sons, New York, 1968.
44. K. Pearson, "The Problem of the Random Walk," *Nature* LXXII, 1905, pp. 294, 342.
45. J.C. Kluyver, "A Local Probability Problem," *Proceedings Section of Science, K. Akad. Van Wet. Te Amsterdam*, VIII, 1905-6, pp. 3441-3450.

46. Lord Rayleigh, "On the Problem of Random Vibrations, and of Random Flights in One, Two, or Three Dimensions," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol. 37, No. 220, 1919, pp. 321-347.
47. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd Edition, John Wiley & Sons, New York, 1971, pp. 32-33.
48. G.N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd Edition, Cambridge University Press, 1980, pp. 419-421.
49. E. Jakeman and P.N. Pusey, "A Model for Non-Rayleigh Sea Echo," *IEEE Transactions on Antennas and Propagation*, Vol. AP-24, No. 6, 1976, pp. 806-814.
50. T.K. Keyes and W.T. Tucker, "The K-distribution," General Electric Internal Report (unpublished), 1997.
51. Reference 52, footnote 7.
52. Reference 52, footnote 18.
53. E. Jakeman and P.N. Pusey, "A Model for Non-Rayleigh Sea Echo," *IEEE Transactions on Antennas and Propagation*, Vol. AP-24, No. 6, pp. 806-814.
54. E. Jakeman and R.J.A. Tough, "Non-Gaussian Models for the Statistics of Scattered Waves," *Advances in Physics*, Vol. 37, No. 5, pp. 471-529.
55. G. Wahba, "Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation," *The Annals of Statistics*, Vol. 3, No. 1, January 1975.
56. P.J. Howard, General Electric Aircraft Engine Division, private communication.
57. M. Schwartz, *Information Transmission, Modulation and Noise*, McGraw-Hill, New York, 1980, pp. 377-382.
58. P.D. Panetta, F.J. Margetan, I. Yalda, and R.B. Thompson, "Ultrasonic Attenuation Measurements in Jet Engine Titanium Alloys," in *Review of Progress in Quantitative NDE*, Vol. 15B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1525-1532, 1996.
59. I. Yalda, F.J. Margetan, and R.B. Thompson, "An Evaluation of the Applicability of Rician Distributions to the Addition of Ultrasonic Flaw Signals and Backscattered Noise," in *Review of Progress in Quantitative NDE*, Vol. 17, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 105-112, 1998.

60. P.D. Panetta, F.J. Margetan, I. Yalda, and R.B. Thompson, "Observation and Interpretation of Microstructurally Induced Fluctuations of Back-Surface Signals and Ultrasonic Attenuation in Titanium Alloys," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 16B, D. O. Thompson and D. E. Chimenti, eds., Plenum Press, NY, pp. 1547-1554, 1997.
61. M. Bilgen and J.H. Rose, "Rough Surface Effects on Incoherent Scattering From Random Volumetric Scatterers: Approximate Analytic Series Solution," *Journal of Acoustical Society of America*, 96, 1994, p. 2856.
62. M. Bilgen and J.H. Rose, "Effects of One-Dimensional Random Rough Surfaces on Ultrasonic Backscatter: Utility of Phase-Screen and Fresnel Approximations," *Journal of Acoustical Society of America*, 96, 1994, p. 2849.
63. M. Bilgen and J.H. Rose, "Focused Ultrasonic Probes and the Effects of Surface Roughness on Material Noise," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 13, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, p. 1769, 1994.
64. M. Bilgen and J.H. Rose, "Acoustic Backscatter From Materials with Rough Surfaces and Finite Microstructures: Theory," *Journal of Acoustical Society of America*, 101, 1997, p. 264.
65. J.H. Rose, M. Bilgen, and P.B. Nagy, "Acoustic Double-Reflection and Transmission at a Rough Water-Solid Interface," *Journal of Acoustical Society of America*, 95, 1994, p. 3242.
66. M. Bilgen and J.H. Rose, "Doubly-Coherent Transmission at Rough Surfaces and Its Implications for Ultrasonic Inspection," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 13B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, p. 1753, 1994.
67. M. Bilgen and J.H. Rose, "Mean and Variance of the Ultrasonic Signal From a Scatterer Beneath a Rough Surface: Theory," *J. Acoust. Soc. Am.* 98, 1995, p. 2217.
68. M. Bilgen and J.H. Rose, "Ultrasonic Signals From "Worse-Case" Hard-Alpha Inclusions Beneath a Random Rough Surface," in *Review of Progress in Quantitative NDE*, Vol. 14B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1837-1844, 1995.
69. P.B. Nagy, J.H. Rose, and M. Bilgen, "Excess Scattering Induced Loss at a Rough Surface Due to Partially Coherent Double-Reflection," in *Review of Progress in Quantitative NDE*, Vol. 14B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1845-1852, 1995.

70. K. Holman, M. Holland, J. Miller, P. Nagy, and J. Rose, "Effective Ultrasonic Transmission Coefficient for Randomly Rough Surfaces," *Journal of Acoustical Society of America*, 100, 1996, p. 832.
71. M. Bilgen and J.H. Rose, "Variance of the Ultrasonic Signal From a Defect Beneath a Rough Surface," in *Review of Progress in Quantitative NDE*, Vol. 15B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1463-1470, 1996.
72. R. Roberts and J.H. Rose, unpublished results.
73. E.J. Nieters, et al., "A Multizone Technique for Billet Inspection," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 14B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 2137-2144, 1995.
74. R. Roberts, unpublished results.
75. M. Bilgen and J.H. Rose, "Acoustic Signal-to-Noise Ratio for Flaws Beneath a Randomly Rough Surface and in the Presence of Microstructure," *Journal of Acoustical Society of America*, 101, 1997, p. 272.
76. J.H. Rose and M. Bilgen, Center for Nondestructive Evaluation, Iowa State University, private communication.
77. C.-P. Chiou, I. Yalda, F.J. Margetan, R.B. Thompson, and D.J. Sturges, "The Use of Ultrasonic Flaw and Noise Models in Designing Titanium Test Blocks," in *Review of Progress in Quantitative NDE*, Vol. 17B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 2069-2076, 1998.
78. L. Brasche and K. Smith, "Engine Titanium Consortium: Status of the Contaminated Billet Study," in *Materials and Process Affordability: Keys to the Future*, H.S. Kliger, B.M. Rasmussen, L. A. Pilato, and T.B. Tolle, eds., SAMPE, Corina, CA, 1998, pp. 1458-1464.
79. C.-P. Chiou, B. Boyd, R.B. Thompson, and J. Oliver, "Geometrical and Ultrasonic Modeling of Hard-Alpha Defect Components in Titanium," in *Review of Progress in Quantitative NDE*, Vol. 18, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1829-1836, 1998.

APPENDIX A—DETAILS OF FLAW SIGNAL MODELING AND VALIDATION

An important aspect of the new methodology is using physical models, to the extent possible, to predict flaw signals under the influence of various inspection and material parameters. This reduces the experimental effort and provides a basis for considering cases not covered by experiment after proper validation has occurred. In the following discussions, the ultrasonic, noise-free flaw signal models and their validations will be described to demonstrate the accuracy and the range of applicability of these models.

A.1 THE AULD-THOMPSON-GRAY FRAMEWORK.

The ultrasonic flaw response modeling approach in this work was derived within the context of Auld's electromechanical reciprocity relation [A-1] as has been demonstrated previously in the derivation of the Thompson-Gray measurement model [A-2]. Auld's reciprocity relation provides a general and rigorous way to relate ultrasonic scattering theory to the absolute level of signals observed at the electrical ports of a transducer. The key in this process is the specification of the wavefields in the vicinity of the flaw, in terms of scattering theory. Within this framework, the Thompson-Gray measurement model describes an immersion inspection process where the flaw is small with respect to the ultrasonic beam width. In this small flaw limit, the prediction of the overall signal is made in terms of the product of a number of factors, which may be thought of as theoretical modules that describe the flaw-scattering response, beam propagation, medium attenuation, and interface transmission effects. In order to make absolute predictions of flaw response, inference of the system response obtained from a separate reference experiment is required. The measurement model is formulated in the frequency domain. For a given inspection geometry and flaw particulars, the absolute values of various radio frequency (RF) flaw waveforms can be predicted to great accuracy by using inverse Fourier transform techniques.

Symbolically, the frequency domain components predicting the results of the inspection process, as predicted by the Thompson-Gray measurement model, can be summarized in the following expression:

$$\begin{bmatrix} \text{Spectral} \\ \text{component of} \\ \text{flaw signal} \end{bmatrix} \propto \begin{bmatrix} \text{probe efficiency} \end{bmatrix} \begin{bmatrix} \text{interface} \\ \text{transmission} \end{bmatrix} \begin{bmatrix} \text{diffraction, focusing,} \\ \text{and other effects} \end{bmatrix} \begin{bmatrix} \text{phase \&} \\ \text{attenuation} \end{bmatrix} [\text{flaw signature}] \quad (\text{A-1})$$

The “diffraction, focusing, and other effects” submodule, which describes the effects of beam propagation, as influenced by part geometry, is calculated based on the Gauss-Hermite beam model [A-3]. The “probe efficiency factor” submodule, used to account for the system efficiency, is deduced from a measured RF reference signal in which backsurface echoes obtained from a fused quartz plate are normally used¹. This reference signal together with other geometry and material parameters, i.e., “interface transmission,” “phase & attenuation,” and “flaw signature,” are the necessary inputs to predict the signal response. After the “spectral component of flaw signal” is evaluated at all frequency components within the transducer

¹ The use of fused quartz as reference samples is attributable to R. S. Gilmore.

bandwidth, the time domain RF signal can then be synthesized via fast Fourier transform techniques. The main advantage of this approach is the reduction of computational effort, while maintaining necessary modeling complexity, through the introduction of appropriate approximations in the various modules. Using the Gauss-Hermite beam model, for example, enables the wave fields to be rapidly calculated in the regions of interest which allows the “diffraction, focusing, and other effects” submodule to be evaluated. In contrast with other numerical methods such as finite element and boundary element, this beam model does not require as much computation time. However, this advantage is gained by using paraxial approximation, which can break down at large angles of incidence or when using very high numerical aperture focused probes. Determining the range of applicability through validation experiments will be discussed in subsequent experiments.

The central part of the ultrasonic model development is the flaw signature, which describes the ultrasonic characteristics of flaws of various types and geometries. This will be discussed in more detail in the following sections. For extended flaws that are not small with respect to the beam, the form of equation A-1 must be modified in various ways, as will also be discussed.

A.2 APPLICATION TO FLAT-BOTTOM HOLES (FBHs)

The FBH is of considerable interest because of its role as an industrial reference standard. Note for example, the first recommendation in section 4.1. The ability to detect FBHs of various sizes in specified situations was a major component of the Titanium Rotating Component Review Team report [A-4], as noted in section 4.1.

To the best knowledge, an exact model for the response of the FBH is not available. In principle, a model could be developed based on numerical approaches such as the finite element or finite difference method. Since such calculations are quite computationally intensive in three-dimensions and the models are intended to be used repeatedly throughout the methodology, as well as other applications such as system design, it was decided that approximate models would run more quickly on a PC or workstation. An accuracy target of 3 dB, based on a consensus evaluation of the reproducibility of experiments in an industrial setting, was established.

A review of the literature indicated that a number of models were available, and examination of their assumptions indicated that each was most useful in different regions of wavelength, beam diameter, or flaw size space. Pioneering work in this area was conducted by Krautkrämer [A-5], who developed a model for the far field response of an FBH. Three more recent models, namely, the method of optimal truncation [A-3], the plane-wave Kirchhoff approximation [A-6], and the finite-beam Kirchhoff approximation, were used in this study along with the plane-wave Kirchhoff approximation, which is closely related to Krautkrämer’s original work. In each case, the effects of the circular-cylindrical surface, created by forming the shanks of the FBH, are neglected and the FBH response is approximated by a penny-shaped crack. A hybrid strategy of combining these three models was then adapted for this situation [A-7].

In modeling the flaw-scattering responses for the cases where the flaw size was smaller than the beam width, a number of traditional approaches predicting a quantity known as the scattering amplitude under the assumption of plane-wave ensonification are readily available. These define

the flaw signature module in equation A-1. One such plane-wave solution, the Method of Optimal Truncation (MOOT) [A-8], is particularly suitable for this FBH study. This method expresses the plane-wave solution in terms of series expansion truncated optimally in the least-squares sense. Its predictions can be considered nearly exact when the flaw is small with respect to the beam. The plane-wave Kirchhoff model is another plane-wave solution (hereafter denoted by PKIR with KIR referring to Kirchhoff [A-6]). In PKIR, the high-frequency Kirchhoff approximation, is used for plane-wave-illuminated elliptical cracks whose local reflectivity is approximated by that of an infinite plane. Simple closed-form solutions can be obtained for such a problem, allowing very rapid evaluations. However, the assumption that the local reflectivity can be approximated by that of an infinite plane (Kirchhoff approximation) limits the applicability of this approximation to cracks whose dimensions are on the order of the ultrasonic wavelength or larger. At longer wavelengths, representation of local reflectivity by that of an infinite plane begins to break down.

For a flaw size comparable to or greater than the beam width, however, the amplitude variation of the incident sonic field over the surface of the flaw becomes significant and must be carefully treated, since plane-wave illumination is no longer a valid approximation. An approach combining the Kirchhoff approximation with numerical integration has been developed for this purpose [A-9]. In this approach, the Kirchhoff assumption along with other boundary conditions are used to simplify Auld's reciprocity relation. This simplification allows the scattered flaw response to be approximated by a numerical integration over the flaw area of the square of the incident displacement field, which can be modeled by the Gauss-Hermite series expansion for the transducer radiation pattern [A-3]. In this case, equation A-1 is no longer valid but must be generalized to include an integral over the flaw surface. This approximation is known as the finite-beam Kirchhoff approximation (FKIR).

In their current implementation, both PKIR and FKIR are generally believed to be applicable to cases of small tilt angles of the planar disk relative to the axis of the sound beam and to flaw-wavelength ratios greater than 1 or 2. The flaw-wavelength ratio is defined as $2\pi a/\lambda$ where a is the flaw radius and λ is the ultrasonic wavelength. The notation ka is sometimes used as a shorthand for this important physical quantity. ka describes how the wavelength scales with respect to the size of the flaw and, hence, whether one is in a regime in which various interferences might occur that would influence the flaw response. Whereas MOOT, valid for all tilt angles, is numerically limited to $2\pi a/\lambda < 10$, FKIR and PKIR apply to small tilt angles and flaws for which $2\pi a/\lambda > 1$ or 2. In addition, both PKIR and MOOT are limited to flaws small with respect to the beam dimensions, whereas FKIR can handle larger flaws. In the strictest sense, MOOT is valid only for circular cracks while PKIR and FKIR can be applied to cracks of elliptical or more complex shape. Table A-1 summarizes these ranges of validity along with their relative computation speeds.

To determine the accuracy of the model predictions, extensive experimental data were taken from a titanium specimen using three different transducers at various angles of incidence and focal depths. This titanium specimen was fabricated from a Ti-6Al-4V ring forging machined into a flat plate. It contained 64 FBHs, 16 each of sizes no. 1 to 4, all normal to the entry (front) surface, and with hole ends at 1" depth below the entry surface. The convention of designating the diameter of an FBH in units of 1/64" (0.4 mm) is still being used. The three transducers were

TABLE A-1. RANGE OF VALIDITY OF VARIOUS APPROXIMATE MODELS FOR FBHS

Model	$2\pi a/\lambda$	Flaw size Beam size	Tilt Angle	Computational Speed
FKIR	> 1 or 2	all	Small	medium
PKIR	> 1 or 2	small	Small	fast
MOOT	< 10	small	All	fast

chosen to be 5-MHz broadband, focused immersion probes with nominal values of 1" diameter and 8" focal length. Independent beam mapping experiments were used to deduce the effective diameter and geometrical focal length of each transducer, as summarized in table A-2.

TABLE A-2. PARAMETERS OF TRANSDUCERS USED

Transducer	Make	Frequency (MHz)	Diameter (inch)	Geometrical focal length (inch)	True focal length (inch)
1	Panametrics V307	5	1	11.14	8.50
2	Panametrics V307	5	1	10.80	8.40
3	Ultran WS100-5-P8	5	1	7.79	6.70
4	GE CRD94-6	10	1.5	9.79	9.25

The geometrical focal lengths of the first two transducers (hereafter referred as transducers no. 1 and no. 2; made by the same vendor) were approximately 11" in water while the third (transducer no. 3), made by another vendor, was about 3" shorter. The true focal length of the transducers also varies but to a lesser extent. Here the geometrical focal length is the focal distance that would be computed by a (high-frequency) ray analysis, whereas the true focal length is the distance to the peak pressure point; a distance which is smaller due to diffraction. Table A-2 also lists the parameters of a 10-MHz broadband, focused probe of 1.5" diameter (transducer no. 4), which was used in the synthetic hard-alpha inclusion study (see section A.3).

Using each of the first three transducers, the FBH sample was scanned ten times following a matrix involving three positions of the focal plane with respect to the flaw plane (transducer focused 1/4" above the flaw, on the flaw, and 1/2" below the flaw) and three angles of incidence (0°, 2.5°, and 5° with respect to the normal) with one repeat of the "normal incidence focused on flaw" case. Scanning step size was 0.010" (10 mils) in both x and y directions. The data obtained are stored, providing

- a basis for computing noise (from data when the beam was well away from any flaws),
- the mean flaw response (average of the maximum response of the 16 nominally identical flaws) used to validate the theory, and
- distributions of flaw response (based on the variabilities of the response of nominally identical flaws).

Further details may be found in reference A-10.

Figure A-1 displays three experimental RF waveforms and the corresponding model predictions by the FKIR for a typical no. 4 FBH response at normal incidence when observed with transducer no. 3 focused on the plane of the FBHs. This illustrates the good agreement that can be achieved in both amplitude and phase for one on-axis and two offset cases. Recall that this is an absolute comparison, with no adjustable parameters, and the overall system efficiency being established by the reference experiment.

In figures A-2 through A-4, comparisons are made between the experiments and predictions of the three individual models for different tilt angles. The experimental data for this series are taken using transducer no. 2, with the beam again focused on the plane of FBHs. The peak-to-peak model predictions, along with mean experimental values (averaged over the 16 nominally identical holes in each of four sizes), are plotted in figure A-2 for the case of normal incidence. It is seen that all models do a reasonable job. Because of the finite-beam effect, a deviation of MOOT and PKIR predictions from experiments at the larger hole sizes, is expected since the approximation of plane-wave ensonification is no longer valid. It is also interesting to see that a “bump” occurring near $ka = 1$ (i.e., near the no. 1 hole size), as predicted by MOOT, is consistent with the experimental data. PKIR and FKIR, on the other hand, do not include a prediction of this bump feature due to the simplification in the Kirchhoff approximation. This bump, as predicted by MOOT, is attributed to a resonance generated by the interaction between Rayleigh waves propagating on the face of the flaw and the flaw edges.

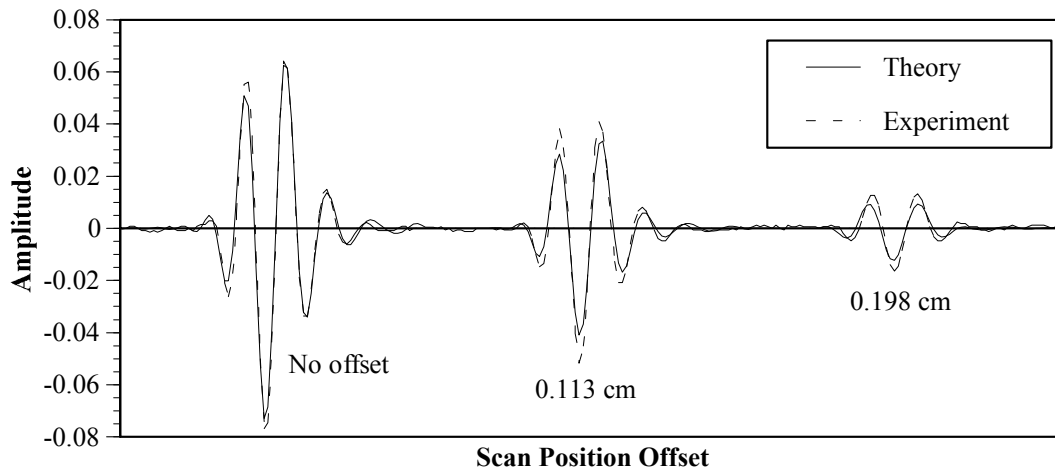


FIGURE A-1. ABSOLUTE AMPLITUDE AND PHASE COMPARISONS BETWEEN MODEL AND EXPERIMENT FOR A TYPICAL NO. 4 FBH AT NORMAL INCIDENCE USING TRANSDUCER NO. 3 FOCUSED ON HOLE

At a 2.5° tilt angle in water (i.e., a tilt angle of approximately 10° in the titanium), model predictions shown in figure A-3 follow a trend similar to the normal incidence cases. Note that the differences between the models have been reduced because of the smaller projected FBH size, and the resonance bump has shifted beyond the region of $ka = 1$. When the tilt angle increases to 5° , leveling off of the response occurs around FBH diameters of 1.5 mm, i.e., 0.059"

or approximately 4/64" (figure A-4). Examination of time domain responses suggests that this leveling off is associated with a separation of the tip-diffracted signals from the leading and trailing edges of the flaw. Note that edge-diffracted waves occurring in FBH examinations will be quantitatively different from those in crack examinations, which are modeled by MOOT. This is because the edges of a FBH have 90° angles and those of a crack have an 180° angle. However, the overall situations are expected to be similar.

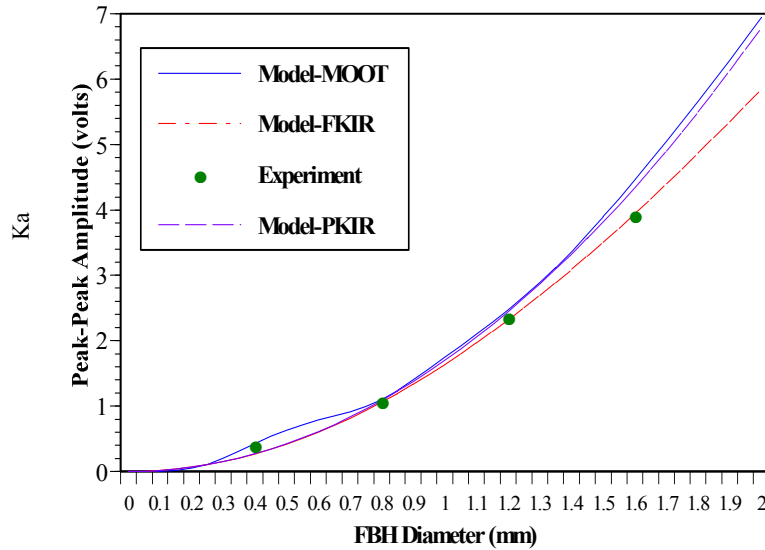


FIGURE A-2. PEAK-TO-PEAK SIGNAL COMPARISONS BETWEEN MODEL AND EXPERIMENT USING TRANSDUCER NO. 2 FOCUSED ON FBHs AT NORMAL INCIDENCE

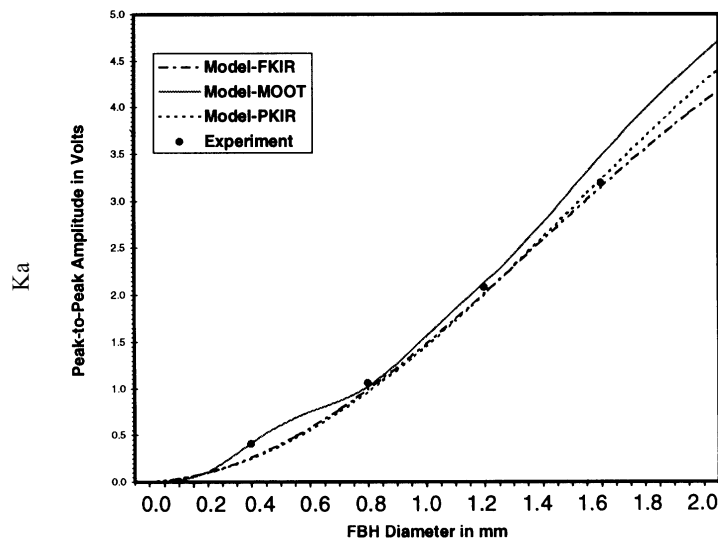


FIGURE A-3. PEAK-TO-PEAK SIGNAL COMPARISONS BETWEEN MODEL AND EXPERIMENT USING TRANSDUCER NO. 2 FOCUSED ON FBHs AND 2.5° TILT ANGLE IN WATER

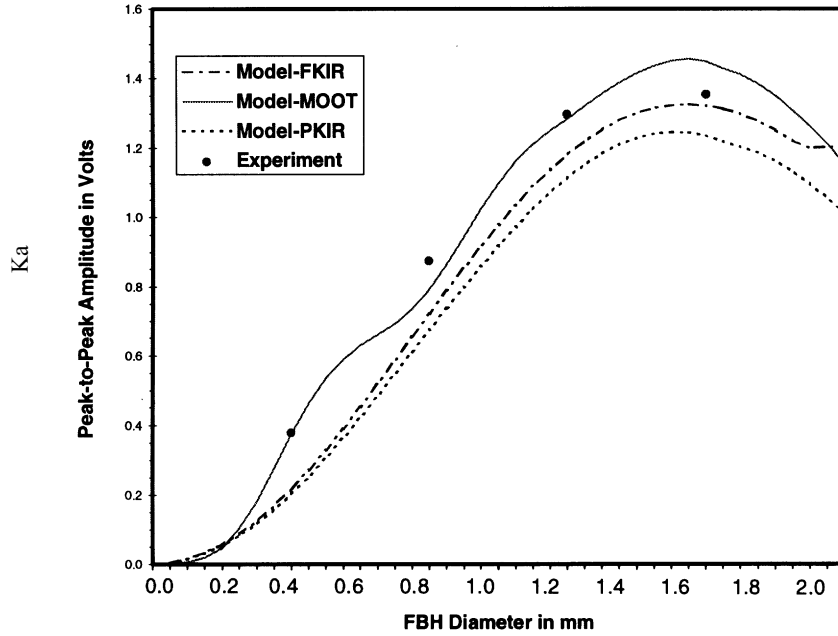


FIGURE A-4. PEAK-PEAK SIGNAL COMPARISONS BETWEEN MODEL AND EXPERIMENT USING TRANSDUCER NO. 2 FOCUSED ON FBHs AND 5° TILT ANGLE IN WATER

From previous theoretical discussions in section A.2, it is noted that MOOT, as an exact plane-wave incidence model, can successfully predict the resonance around $ka = 1$, which is accurate for smaller FBH sizes where the finite-beam effect is not significant. However, the accuracy drops as the FBH size increases. In contrast, FKIR is good at handling finite-beam effects for large FBH sizes (high $2\pi a/\lambda$) but is less accurate in the smaller size region (low $2\pi a/\lambda$). On the other hand, PKIR, as an approximated plane-wave model, performs similarly to MOOT when the FBH size becomes larger and has the same characteristics as FKIR in the small size region. Each of these theoretical expectations is confirmed by the data shown in figures A-2 through A-4.

For this purpose, however, a single model is needed that gives accurate predictions over the entire range of flaw sizes. Given that MOOT is best for small flaws and FKIR is best for large flaws, a means is desired to smoothly join these two. This could be accomplished using a hypothesized simple hybrid relationship that is suitable for a wide range of FBH sizes by combining the strength of each of the three models:

$$\text{Hybrid Model FBH response} = \frac{\text{MOOT} \times \text{FKIR}}{\text{PKIR}} \quad (\text{A-2})$$

In the small flaw limit, the response will be predicted by MOOT, since $\text{PKIR} \sim \text{FKIR}$. In the large flaw limit, the response will essentially be predicted by FKIR, since $\text{MOOT} \sim \text{PKIR}$.

The usefulness of equation A-2 can be seen from figure A-5, where a single hybrid model curve is considered to be a better match with experimental data for the entire range of FBH sizes than the three individual model curves, each of which is only accurate over a portion of the size range.

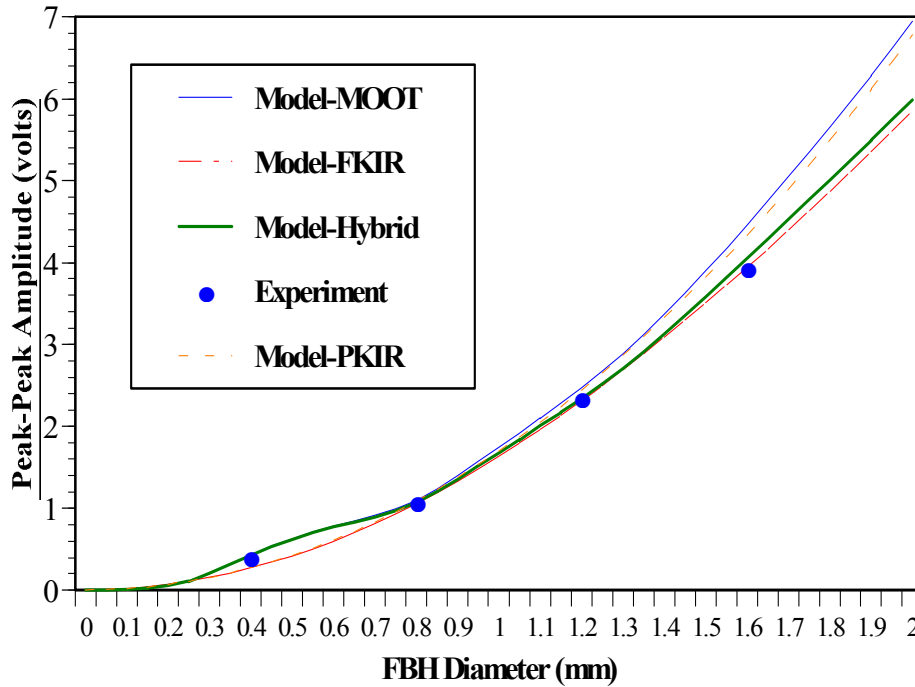


FIGURE A-5. PEAK-TO-PEAK SIGNAL COMPARISONS BETWEEN HYBRID MODEL PREDICTION AND EXPERIMENT USING TRANSDUCER NO. 2 AT NORMAL INCIDENCE, FOCUSED ON FBHs

Tables A-3 through A-5 present a comparison of the hybrid model predictions to mean experimental data for each of the 36 cases examined (4 flaw sizes x 3 focal depths x 3 tilt angles) for the three transducers used. It can be seen, based on a better than 40% agreement between the EXPT and MODL responses, that, in general, the criteria of 3-dB accuracy has been realized. The one exception is for the no. 4 FBH, viewed with transducer no. 3, and focused 0.5 inch above the flaw at a tilt of 2.5° in water. Even here, the disagreement is only 3.6 dB.

TABLE A-3. COMPARISON OF AVERAGED PEAK-TO-PEAK FBH SIGNALS AND THE CORRESPONDING STANDARD DEVIATIONS (DENOTED BY EXPT) TO MODEL PREDICTIONS FOR ALL CASES USING TRANSDUCER NO. 1

Focal Depth	Incident Angle	Flat-Bottom Hole Size				
		No. 4	No. 3	No. 2	No. 1	
1"	0	3824 105	2280 78	1049 31	399 40	EXPT
		[3888]	[2279]	[1043]	[266]	FKIR
		(4551)	(2525)	(1141)	(424)	MOOT
		{4492}	{2527}	{1123}	{281}	PKIR
		<3939>	<2277>	<1060>	<401>	MODL
1"	-2.5	3054 96	2014 80	1007 48	378 34	
		[3195]	[2048]	[1002]	[266]	
		(3588)	(2192)	(1064)	(411)	
		{3546}	{2199}	{1047}	{272}	
		<3233>	<2041>	<1018>	<402>	
1"	5	1446 70	1347 48	883 28	365 33	
		[1340]	[1184]	[718]	[216]	
		(1552)	(1348)	(840)	(367)	
		{1561}	{1381}	{823}	{245}	
		<1332>	<1156>	<733>	<324>	
1.25"	0	3411 120	2069 85	956 33	343 34	
		[3541]	[2086]	[963]	[246]	
		(4277)	(2356)	(1074)	(391)	
		{4217}	{2372}	{1054}	{264}	
		<3549>	<2072>	<994>	<364>	
1.25"	-2.5	2731 109	1812 73	899 40	322 34	
		[3053]	[1954]	[955]	[253]	
		(3394)	(2076)	(1013)	(382)	
		{3394}	{2091}	{991}	{257}	
		<3053>	<1940>	<976>	<376>	
1.25"	5	1472 55	1351 53	853 27	337 32	
		[1401]	[1191]	[707]	[210]	
		(1598)	(1358)	(830)	(356)	
		{1612}	{1364}	{808}	{238}	
		<1389>	<1186>	<726>	<314>	
0.5"	0	2905 77	1713 68	782 33	298 31	
		[3129]	[1811]	[821]	[208]	
		(3557)	(1981)	(898)	(331)	
		{3482}	{1959}	{871}	{218}	
		<3196>	<1831>	<846>	<316>	
0.5"	-2.5	2243 74	1431 52	715 35	275 27	
		[2458]	[1567]	[762]	[201]	
		(2695)	(1658)	(807)	(306)	
		{2652}	{1657}	{793}	{207}	
		<2498>	<1568>	<775>	<297>	
0.5"	5	1015 54	894 38	583 27	251 25	
		[964]	[855]	[518]	[155]	
		(1078)	(957)	(598)	(265)	
		{1078}	{980}	{597}	{178}	
		<964>	<835>	<519>	<231>	
1"	0	3827 106	2274 90	1046 35	398 40	
		<3939>	<2277>	<1060>	<401>	

Note: FKIR (numbers within []) represents the finite-beam Kirchhoff model, MOOT (numbers with ()) is the method of optimal truncations, PKIR (numbers within {}) denotes the plane-wave Kirchhoff model and MODL (numbers within <>) designates the hybrid model prediction as computed by using equation A-2. Measurement units are focal depth – inch; incident angle – degree; FBH size - no. 1 = 1/64", no. 2 = 2/64", etc.; FBH amplitude - millivolts.

TABLE A-4. COMPARISON OF AVERAGED PEAK-TO-PEAK FBH SIGNALS AND THE CORRESPONDING STANDARD DEVIATIONS (DENOTED BY EXPT) TO MODEL PREDICTIONS FOR ALL CASES USING TRANSDUCER NO. 2

Focal Depth	Incident Angle	Flat-Bottom Hole Size				
		No. 4	No. 3	No. 2	No. 1	
1"	0	3940 113	2368 93	1086 40	410 41	EXPT
		[3888]	[2279]	[1043]	[266]	FKIR
		(4620)	(2566)	(1157)	(422)	MOOT
		{4535}	{2551}	{1134}	{283}	PKIR
		<3961>	<2292>	<1064>	<397>	MODL
1"	-2.5	3189 96	2077 66	1050 50	401 35	
		[3148]	[2023]	[991]	[263]	
		(3596)	(2207)	(1070)	(410)	
		{3544}	{2210}	{1056}	{276}	
		<3194>	<2020>	<1004>	<391>	
1"	5	1353 73	1296 51	874 29	377 34	
		[1322]	[1182]	[722]	[218]	
		(1529)	(1352)	(836)	(369)	
		{1528}	{1365}	{819}	{244}	
		<1323>	<1171>	<737>	<330>	
1.25"	0	3381 99	2074 83	966 43	345 35	
		[3513]	[2073]	[955]	[244]	
		(4229)	(2347)	(1072)	(394)	
		{4189}	{2356}	{1047}	{262}	
		<3547>	<2065>	<978>	<367>	
1.25"	-2.5	2856 94	1899 75	949 44	321 33	
		[2977]	[1916]	[941]	[250]	
		(3397)	(2082)	(1016)	(386)	
		{3341}	{2070}	{984}	{256}	
		<3027>	<1927>	<972>	<377>	
1.25"	5	1343 57	1300 55	849 27	350 35	
		[1390]	[1197]	[716]	[213]	
		(1603)	(1374)	(832)	(360)	
		{1609}	{1393}	{819}	{240}	
		<1385>	<1181>	<727>	<320>	
0.5"	0	2906 85	1709 53	777 29	301 32	
		[3047]	[1766]	[802]	[203]	
		(3410)	(1894)	(854)	(321)	
		{3356}	{1888}	{839}	{210}	
		<3096>	<1772>	<816>	<310>	
0.5"	-2.5	2264 85	1433 50	718 41	255 26	
		[2372]	[1515]	[739]	[196]	
		(2647)	(1624)	(787)	(310)	
		{2605}	{1622}	{774}	{202}	
		<2410>	<1517>	<751>	<301>	
0.5"	5	952 48	865 34	576 22	235 27	
		[918]	[818]	[499]	[150]	
		(1042)	(922)	(581)	(258)	
		{1045}	{943}	{569}	{170}	
		<915>	<800>	<510>	<228>	
1"	0	3928 103	2345 90	1083 36	412 41	
		<3961>	<2292>	<1064>	<397>	

Note: FKIR (numbers within []) represents the finite-beam Kirchhoff model, MOOT (numbers with ()) is the method of optimal truncations, PKIR (numbers within {}) denotes the plane-wave Kirchhoff model and MODL (numbers within <>) designates the hybrid model prediction as computed by using equation A-2. Measurement units are focal depth – inch; incident angle – degree; FBH size - no. 1 = 1/64", no. 2 = 2/64", etc.; FBH amplitude - millivolts.

TABLE A-5. COMPARISON OF AVERAGED PEAK-TO-PEAK FBH SIGNALS AND THE CORRESPONDING STANDARD DEVIATIONS (DENOTED BY EXPT) TO MODEL PREDICTIONS FOR ALL CASES USING TRANSDUCER NO. 3

Focal Depth	Incident Angle	Flat-Bottom Hole Size				
		No. 4	No. 3	No. 2	No. 1	
1"	0	7899 195	4808 130	2243 78	763 72	EXPT
		[7712]	[4684]	[2201]	[569]	FKIR
		(9001)	(5062)	(2294)	(883)	MOOT
		{8805}	{4953}	{2201}	{550}	PKIR
		<7883>	<4787>	<2294>	<913>	MODL
1"	-2.5	4286 127	3306 110	1896 81	700 46	
		[5724]	[3840]	[1949]	[529]	
		(6943)	(4366)	(2141)	(846)	
		{6845}	{4329}	{2088}	{548}	
		<5806>	<3873>	<1998>	<817>	
1"	5	2989 126	2641 94	1729 50	647 57	
		[2223]	[2032]	[1310]	[410]	
		(2508)	(2392)	(1533)	(743)	
		{2521}	{2433}	{1549}	{475}	
		<2211>	<1998>	<1296>	<641>	
1.25"	0	5930 137	3755 152	1786 57	597 56	
		[5382]	[3315]	[1586]	[415]	
		(6732)	(3757)	(1699)	(638)	
		{6665}	{3749}	{1666}	{417}	
		<6741>	<3792>	<1617>	<635>	
1.25"	-2.5	3910 108	3124 89	1768 80	647 49	
		[4316]	[2813]	[1417]	[385]	
		(5247)	(3242)	(1580)	(619)	
		{5207}	{3247}	{1551}	{405}	
		<4349>	<2809>	<1443>	<588>	
1.25"	5	2683 98	2353 99	1540 46	560 60	
		[2123]	[1780]	[1088]	[334]	
		(2379)	(2125)	(1318)	(608)	
		(2400)	(2165)	(1304)	(387)	
		<2104>	<1747>	<1100>	<525>	
0.5"	0	4558 109	2659 92	1234 39	431 48	
		[4842]	[2865]	[1321]	[337]	
		(5544)	(3070)	(1366)	(527)	
		{5530}	{3111}	{1382}	{346}	
		<4854>	<2827>	<1306>	<513>	
0.5"	-2.5	2365 68	1716 54	965 44	365 30	
		[3577]	[2308]	[1142]	[305]	
		(3862)	(2402)	(1187)	(489)	
		{3844}	{2437}	{1177}	{309}	
		<3594>	<2275>	<1152>	<483>	
0.5"	5	1641 73	1378 41	882 33	349 37	
		[1301]	[1127]	[703]	[215]	
		(1303)	(1275)	(827)	(377)	
		{1310}	{1295}	{819}	{249}	
		<1294>	<1110>	<710>	<326>	
1"	0	7962 173	4847 172	2280 67	764 69	
		<7883>	<4787>	<2294>	<913>	

Note: FKIR (numbers within []) represents the finite-beam Kirchhoff model, MOOT (numbers with ()) is the method of optimal truncations, PKIR (numbers within {}) denotes the plane-wave Kirchhoff model and MODL (numbers within <>) designates the hybrid model prediction as computed by using equation A-2. Measurement units are focal depth – inch; incident angle – degree; FBH size - no. 1 = 1/64", no. 2 = 2/64", etc.; FBH amplitude - millivolts.

A.3 APPLICATION TO SYNTHETIC HARD-ALPHA INCLUSIONS (SHAs).

The ultrasonic modeling of inclusion signals follows the same Auld-Thompson-Gray approach framework as the previous FBH models. In this work, two extensions of the existing models were developed. The effort first focused on inclusions of cylindrical shape at normal incidence [A-10] and then extended it to arbitrary flaw morphology at oblique incidence [A-11]. In both extensions the paraxial approximation is utilized in the beam model through the Gauss-Hermite series expansions, a feature that is common with the FBH models.

To reduce the modeling complexity and the computation time, various approximations are again necessary given the constraints of the memory and processing capacity of the available computing platforms. In modeling, weak scatterers such as the hard-alpha inclusions, i.e., the Born approximation, has been extensively studied in the literature (see for example [A-12 and A-13]). This approximation simplifies the modeling effort by replacing the unknown scattered field quantities that are in the kernel of various integral representations with their incident counterparts on the surface or within the volume of the flaw. In our new formulation, the Auld's surface reciprocity relationship was first converted into a volumetric form via Gauss' theorem. The paraxial and Born approximations were used to reduce the new form to a three-dimensional volumetric integral that involved the square of incident displacement field and the inclusion properties. Numerical integration was then conducted to carry out the computation. The symbolic form for this model can be expressed as

$$\left[\begin{array}{l} \text{spectral} \\ \text{component} \\ \text{of flaw signal} \end{array} \right] \propto \iiint_{\text{flaw vol.}} \left[\begin{array}{l} \text{incident} \\ \text{displacement field} \end{array} \right]^2 f(\text{material and acoustic property}) dV \quad (\text{A-3})$$

In the ordinary context of the Born approximation, both the density and wave speed are assumed to be very close to those of the host materials, i.e., the flaw is assumed to possess a weak (small) impedance mismatch. Intuitively, this weak impedance assumption is consistent with the Born approximation that replaces the wavefield quantities within the inclusion by the incident field counterparts. However, from a comparative study with the high-frequency Kirchhoff approximation [A-14], both approximations (Born & Kirchhoff) have been shown to be equivalent in modeling the leading specular (front surface) responses. Thus, it is unnecessary to impose this weak impedance assumption. The models with and without the weak impedance assumption are referred below as “weak Born model” and “normal Born model.” As will be shown later, the normal Born model offers better accuracy in comparison with the experimental data for inclusions of higher impedance. The full range of applicability of these two models is yet to be determined.

The third model is an ad hoc surface formulation that represents a natural extension to the previous high-frequency finite-beam Kirchhoff model (FKIR) for FBH. In that approach, the Kirchhoff approximation along with other boundary conditions are used to simplify the Auld's reciprocity relation. This simplification allows the scattered flaw response to be approximated by a two-dimensional numerical integration over the flaw surface of the square of the incident field. In the present case of scattering from an inclusion, the Kirchhoff assumption is replaced

by ad hoc boundary conditions on the inclusion-host interface. In essence, one replaces a unity reflection coefficient (FBH) with one determined by the relative acoustic impedances of the SHA and host material. The resulting formulation closely resembles that of previous Kirchhoff models for FBHs, except that an additional surface integration on the inclusion back wall, and a factor for time delay between front and back wall are needed. The symbolic expression of the ad hoc model is given by

$$\begin{bmatrix} \text{spectral} \\ \text{component of} \\ \text{flaw signal} \end{bmatrix} \propto \iint_{\text{front face}} \begin{bmatrix} \text{incident} \\ \text{displacement} \\ \text{field} \end{bmatrix}^2 dA + [\text{delay factor}] \iint_{\text{back face}} \begin{bmatrix} \text{incident} \\ \text{displacement} \\ \text{field} \end{bmatrix}^2 dA \quad (\text{A-4})$$

The ad hoc surface model provides a simple and fast tool for predicting the ultrasonic inclusion responses, especially when numerous C-scan simulations are required. Since this model approximates the scattered field by reflected wave components only, larger modeling errors are expected when other types of responses are present and specularly reflected waves are no longer the dominant components. This situation occurs, for example, when the ultrasonic beam is at a large angle to the inclusion axis and the diffracted wave components emitted from the edges of the cylindrical inclusion become significant.

The two new Born volumetric models are superior to the ad hoc inclusion model in that they are readily applied to relatively weak inclusions of arbitrary morphology as long as the geometry, material, and acoustic properties (or their close estimates) are available everywhere within the inclusion boundaries. The trade-off is that the three-dimensional volumetric integration will consume extra computing time compared to the ad hoc model's double, two-dimensional surface integration.

The ultrasonic models described above have been validated using experimental data obtained from two rectangular Ti-6Al-4V ring forging blocks containing synthetic hard-alpha (SHA) inclusions of cylindrical shape. These inclusions were specially prepared [A-15] to accommodate prescribed nitrogen and oxygen concentrations for the purpose of simulating the real, hard-alpha defects. One block (hereafter labeled as block no. 1) has a total of 64 SHAs of 2.71 wt.% nitrogen and 0.387 wt.% oxygen ranging from no. 2 to 5 sizes with 16 of each size. The other (block no. 2) contains 32 SHAs of 5.88 wt.% nitrogen and 0.465 wt.% oxygen with eight of each size. Here the authors have again adopted the flat-bottomed hole convention that size no. 1 represents the actual diameter in units of 1/64", etc. As was shown in table A-1, a new 10-MHz broadband transducer (transducer no. 4) was added to test the effects of higher frequency and resolution. It has a diameter of 1.5" with a measured true focal length of 9.3" and an estimated geometrical focal length of 9.8". Experiments were conducted in the same way as the FBH experiments.

The accuracy of signal modeling, illustrated in figure A-6, compares the RF waveforms from experimental measurement with those from the two model predictions for a typical no. 5 SHA in block no. 1. The data were obtained with transducer no. 2 focused on the SHAs circular end at normal incidence. This shows that the absolute amplitude and phase predicted by both models are in good agreement with the experimental data. Note that, in the Born model prediction, a

slight extra time delay is also seen for the phase-reversed back wall echo. This was expected since, in the Born model, the incident wave was employed in the kernel of the integral, an approximation which does not account for the wave speed difference between the inclusion and the host media. The overall peak-to-peak amplitude comparisons are summarized in table A-6 for SHAs sizes no. 4 and 5. Transducer no. 2 was used, focused on SHAs at normal incidence using block no. 1. The experimental amplitudes listed are the average of six representative SHAs of each size, leaving out the maximum and minimum ones as outliers. One can observe that the agreements are even better than seen in figure A-6, to within 7% relative error, because the SHA signal fluctuations due to the microstructure change are smoothed out by the averaging.

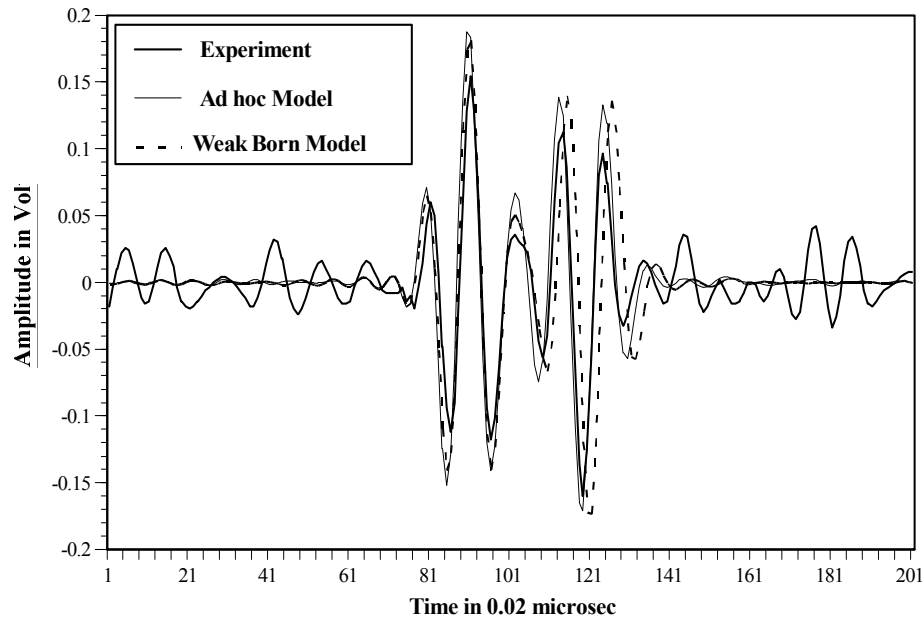


FIGURE A-6. ABSOLUTE AMPLITUDE AND PHASE COMPARISONS BETWEEN MODEL AND EXPERIMENT FOR A TYPICAL NO. 5 SHA IN BLOCK NO. 1 USING THE 5-MHz TRANSDUCER NO. 2 FOCUSED ON THE SHA END AT NORMAL INCIDENCE

TABLE A-6. PEAK-TO-PEAK SIGNAL AMPLITUDE COMPARISONS BETWEEN MODEL AND EXPERIMENT USING BLOCK NO. 1 AND TRANSDUCER NO. 2 AT NORMAL INCIDENCE, FOCUSED ON SHAs

SHA Size	Experiment	Weak Born Model	Ad hoc Model (mV)
No. 5	337	355	359
No. 4	247	234	241

Figure A-7 plots the peak-to-peak amplitudes of averaged experimental data and model predictions vs. various SHA sizes. In this case, transducer no. 2 was operated at normal incidence and focused on SHAs at 1" depth in block no. 2. Good agreement is clearly observed between the models and experiment. The only exception is for size no. 2, for which the model predictions are actually compared with experimental data on the noise floor which obscured the flaw response. In the corresponding C-scan data from which the peak-peak SHA amplitudes

were extracted, none of the No. 2 SHA signals can be identified above the noise level. Both models have correctly confirmed this by predicting lower peak amplitudes. The improvement in amplitude accuracy of the normal Born model over the weak Born model is also significant. The waviness at the lower ends of the model curves is probably due to interference effects between the inclusion front and back wall echoes, which can superimpose constructively or destructively, depending on the specific inclusion sizes in that region.

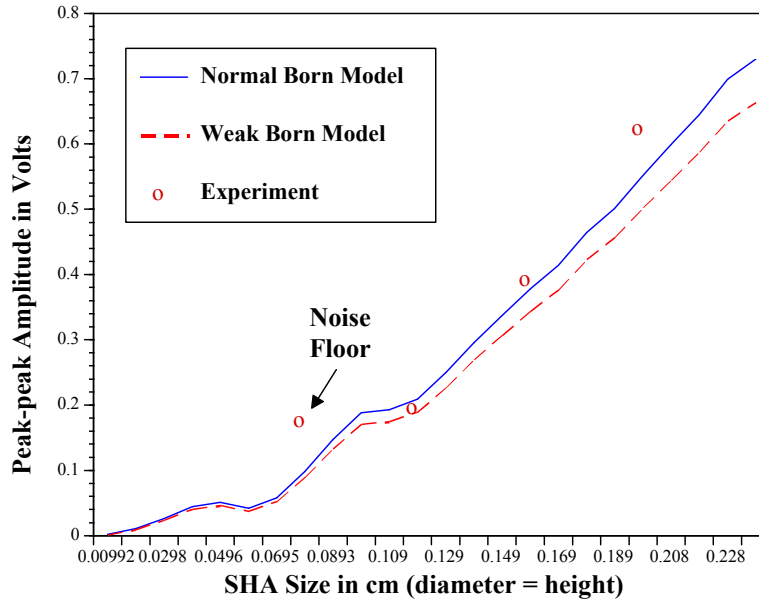


FIGURE A-7. PEAK-TO-PEAK SIGNAL COMPARISONS BETWEEN TWO MODELS AND EXPERIMENT USING TRANSDUCER NO. 2 AT NORMAL INCIDENCE, FOCUSED ON SHAs AT 1" DEPTH IN BLOCK NO. 2

The summary of model predictions and corresponding experimental results are given in tables A-7 and A-8 for all cases at three focal depths and three tilt angles, using transducers no. 2 and 4, respectively. As noted above, the 10-MHz transducer no. 4 was employed here for a comparative study to see how its finer scan resolution would influence the ultrasonic results and the succeeding POTD evaluation. For this study, the step size for transducer no. 2 was set at 0.010" (10 mils) just as in the FBH scans but was reduced to 0.005" (5 mils) in both the x and y directions for transducer no. 4. This was done because of the smaller, diffraction-limited spot site at the higher frequency. Overall, the model predictions show good agreement with experimental data although they underestimate the response in most cases. Since the higher impedance mismatch between the SHA and host titanium is beyond the Born model limit, this underestimate situation is as expected and has been confirmed by a separate comparison between the Born model and an exact solution pertaining to spherical inclusions [A-16].

Also note that the comparisons between the model and experiment are less significant for cases having missed data and are no longer valid for cases where experimental data recorded were actually the noise floor values (indicated respectively by M or N in tables A-7 and A-8). Excluding these cases, the desired 3 dB agreement is observed for all but two cases in table A-7 (hole no. 5, 5° incidence, 1" and 1.25" focal depths). For these, the disagreement is only slightly

worse, having values of 3.7 and 3.6 dB, respectively. Similar comments apply to the higher frequency measurements reported in table A-8. The two cases (hole no. 5, 5° incident angle, 1.25" and 0.5" focal depths) have deviations of 5.0 and 3.9 dB respectively. The fact that those four disagreements all occur for the large SHA (no. 5) and greatest tilt (5° in water or ~20° in the solid) should be the topic of further investigation. However, the fact that the required agreement of 3 dB was observed in all other cases indicates that the primary objective of this phase of the modeling was achieved.

Naturally occurring SHAs often tend to be elongated along the axis of the billet and tend to be ensonified from the side in a normal billet inspection. To test the model's capacity in this geometry, the SHAs were viewed from the side after the transducer was moved to ensonify the edge of the block. Figure A-8 compares the RF waveforms from an experimental measurement with normal Born model predictions for a typical no. 5 SHA in block no. 2 with the beam focused on the SHA's cylindrical side at normal incidence using transducer no. 4. Good agreement is seen in both the absolute amplitude and phase between the model and experiment. The slight time delay of the phase-reversed back wall echo comes from using the incident wave throughout the inclusion body, in the representation integral. As noted above, this does not include the wave speed difference between the inclusion and the host media. In figure A-9, the peak-to-peak amplitudes of all no. 5 SHAs in block no. 2, illuminated from the side using transducer no. 4, are compared with the two Born model predictions. Averaging all eight SHA amplitudes has evidently stabilized the signal fluctuations resulting from microstructure change. The normal Born model is again shown to be superior to the weak Born model by predicting an amplitude closer to the experimental data. By excluding the second SHA amplitude as an outlier, the agreement can be made to within 4% error between the experimental average and the normal Born model. Since the block has not been sectioned, one cannot say why this one inclusion produces a response so much greater than the others. Such a response could be caused by disbonding along the cylindrical side of the inclusion, but there is no evidence that such disbonding has occurred.

A.4 APPLICATION TO NATURALLY OCCURRING HARD-ALPHA INCLUSIONS.

Models have been developed for naturally occurring hard-alpha inclusions, whose response is dominated by small pores. A detailed discussion of these results will appear upon the completion of the Contaminated Billet Study (see section 10.2). Preliminary conclusions, based on two flaws, show that the models agree with the experiment to within ± 6 dB, with possible sources of error, including the reconstruction of the flaws from a sequence of micrographs, the experimental measurements, and the accuracy of the models [A-17].

TABLE A-7. COMPARISON OF AVERAGED PEAK-TO-PEAK SHA SIGNALS AND THE CORRESPONDING STANDARD DEVIATIONS (DENOTED BY EXPT) TO MODEL PREDICTIONS FOR ALL CASES USING TRANSDUCER NO. 2 AND BLOCK NO. 2

Focal Depth	Incident Angle	Synthetic Hard-Alpha Inclusion Size				
		No. 5	No. 4	No. 2	No. 2	
1"	0	624 118	393 31	246 54M	177 38N	EXPT
		[542]	[370]	[244]	[80]	ADHC
		<552>	<378>	<209>	<98>	NMBN
1	-2.5	466 86	322 26	197 24	138 28M	
		<403>	<310>	<194>	<91>	
1	5	201 32	169 21	150 19	91 20	
		<131>	<146>	<144>	<69>	
1.25	0	526 95	337 26	228 43M	151 24N	
		[484]	[321]	[218]	[67]	
		<512>	<347>	<203>	<82>	
1.25"	-2.5	427 65	284 23	184 23	106 17M	
		<388>	<294>	<185>	<78>	
1.25"	5	210 32	168 17	142 18	85 20	
		<139>	<154>	<146>	<66>	
0.5"	0	494 97	292 34	181 50M	155 36N	
		[447]	[300]	[187]	[61]	
		<442>	<293>	<165>	<79>	
0.5"	-2.5	353 68	244 16	137 20	112 16M	
		<300>	<234>	<144>	<71>	
0.5"	5	130 23	112 17	101 12	60 9	
		<93>	<102>	<102>	<50>	
1"	0	630 120	391 31	244 55	177 33N	
		[542]	[370]	[244]	[80]	
		<552>	<378>	<209>	<98>	

Note: Letter N associated with some cases denotes noise floor in which no SHA signal can be identified, while letter M denotes cases when some SHA signals are missed from total eight SHAs of that size. ADHC (numbers within []) represents the ad hoc model, only used at normal incidence, and NMBN (numbers within <>) denotes the normal Born model. Measurement units are focal depth – inch; incident angle – degree; SHA size - no. 2 = 2/64", no. 3 = 3/64", etc.; SHA amplitude - millivolts.

TABLE A-8. COMPARISON OF AVERAGED PEAK-TO-PEAK SHA SIGNALS AND THE CORRESPONDING STANDARD DEVIATIONS (DENOTED BY EXPT) TO MODEL PREDICTIONS FOR ALL CASES USING TRANSDUCER NO. 4 AND BLOCK NO. 2

Focal Depth	Incident Angle	Synthetic Hard-Alpha Inclusion Size				
		No. 5	No. 4	No. 2	No. 2	
1"	0	364 34	252 15	164 32	90 24M	EXPT
		[285]	[215]	[147]	[70]	ADHC
		<304>	<245>	<161>	<77>	NMBN
1	-2.5	163 22	134 14	97 9	69 18M	
		<153>	<143>	<115>	<66>	
1"	5	68 12	56 6	47 8	43 9	
		<41>	<40>	<41>	<40>	
1.25"	0	219 22	156 9	101 17	63 10M	
		[176]	[131]	[88]	44]	
		<208>	<157>	<97>	<45>	
1.25"	-2.5	140 16	110 10	74 8	53 13M	
		<136>	<113>	<82>	<45>	
1.25"	5	73 14	59 6	45 7	41 10	
		<41>	<39>	<38>	<34>	
0.5"	0	195 22	125 67	80 18	47 15M	
		[154]	[112]	[70]	[35]	
		<150>	<112>	<71>	<35>	
0.5"	-2.5	91 17	74 8	50 5	35 8M	
		<85>	<71>	<53>	<29>	
0.5"	5	25 4	22 2	19 2	17 3	
		<16>	<17>	<19>	<15>	
1"	0	332 33	231 18	149 31	92 22M	
		[285]	[215]	[147]	[70]	
		<304>	<245>	<161>	<77>	

Note: Letter N associated with some cases denotes noise floor in which no SHA signal can be identified, while letter M denotes cases when some SHA signals are missed from total eight SHAs of that size. ADHC (numbers within []) represents the ad hoc model and NMBN (numbers within <>) denotes the normal Born model. Measurements units are focal depth – inch; incident angle – degree; SHA size - no. 2 = 2/64", no. 3 = 3/64", etc.; SHA amplitude - millivolts.

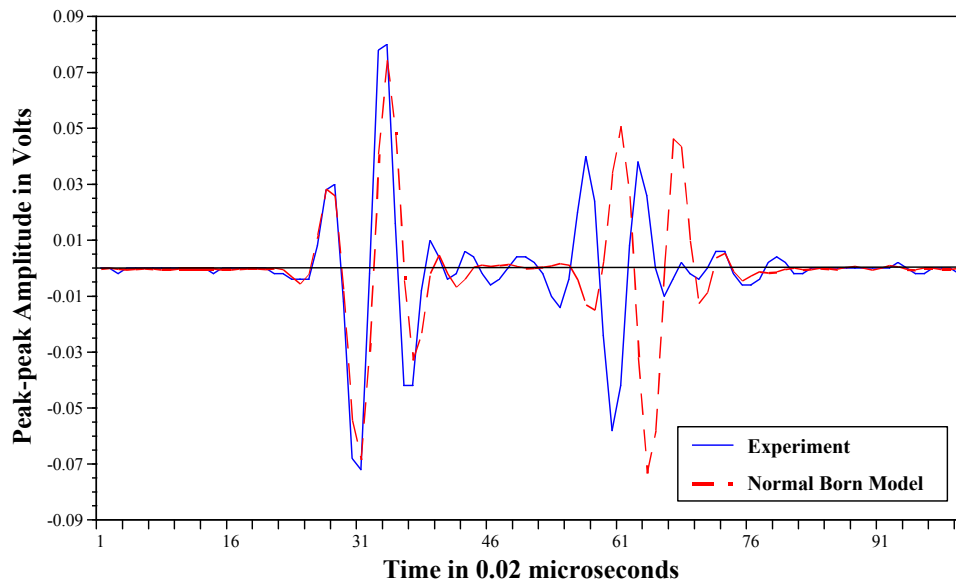


FIGURE A-8. ABSOLUTE AMPLITUDE AND PHASE COMPARISONS BETWEEN MODEL AND EXPERIMENT USING TRANSDUCER NO. 4 AT NORMAL INCIDENCE FOCUSED ON A TYPICAL NO. 5 SHA IN SIDE-ON POSITION IN BLOCK NO. 2

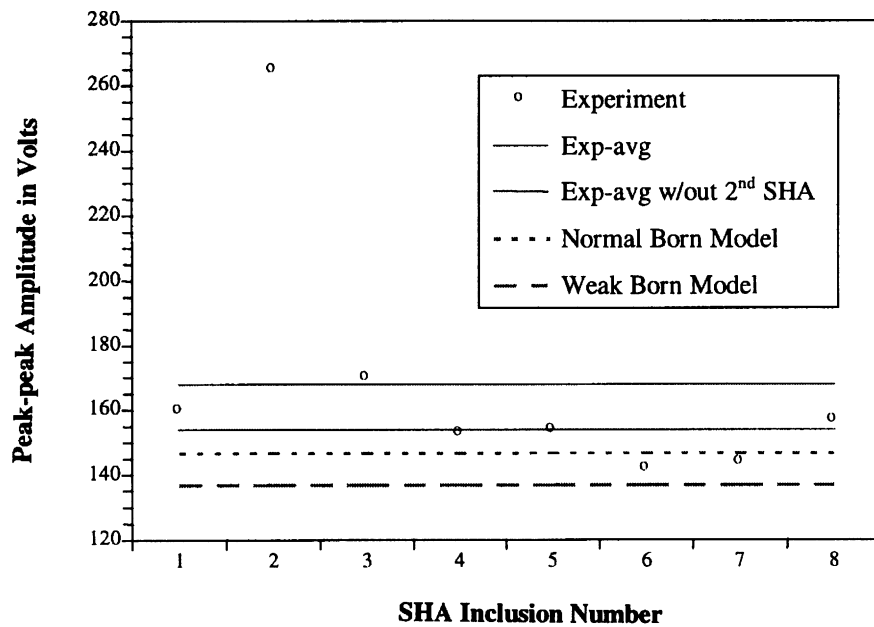


FIGURE A-9. PEAK-TO-PEAK AMPLITUDES OBSERVED WITH TRANSDUCER NO. 4 AT NORMAL INCIDENCE FOR EIGHT NOMINAL NO. 5 SHAs IN BLOCK NO. 2 AT 1.625" DEPTH, COMPARED WITH TWO BORN MODEL PREDICTIONS

A-5 REFERENCE.

- A-1. B.A. Auld, "General Electromechanical Reciprocity Relations Applied to the Calculation of Elastic Wave Scattering Coefficients," *Wave Motion* 1, 1979, pp. 3-10.
- A-2. R.B. Thompson and T.A. Gray, "A Model Relating Ultrasonic Scattering Measurements Through Liquid-Solid Interfaces to Unbounded Medium Scattering Amplitudes," *Journal of Acoustical Society of America*, 74 (4), 1983, pp. 1279-1290.
- A-3. B.P. Newberry, R.B. Thompson, and E.F. Lopez, "Development and Comparison of Beam Models for Two-Media Ultrasonic Inspection," in *Review of Progress in Quantitative NDE*, Vol. 6A, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 639-647, 1987.
- A-4. J.G. Costa, R.E. Gonzalez, R.E. Guyotte, D.P. Salvano, T. Swift, and R.J. Koenig (Chair), "Titanium Rotating Components Review Team Report," Federal Aviation Administration, Aircraft Certification Service, Engine and Propeller Directorate, Boston, MA, December 14, 1990.
- A-5. J. & H. Krautkrämer, "Ultrasonic Testing of Materials," Springer-Verlag, 4th Edition, 1990.
- A-6. L. Adler and J.D. Achenbach, "Elastic Wave Scattering by Elliptical Cracks: Theory and Experiment," *J. Nondestr. Eval.* 1, 1990, p. 87.
- A-7. C.-P. Chiou, F.J. Margetan, and R.B. Thompson, "Ultrasonic Signal Characterizations of Flat-Bottom Holes in Titanium Alloys: Experiment and Theory," in *Review of Progress in Quantitative NDE*, Vol. 14B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 2121-2128, 1995.
- A-8. J.L. Opsal, "Theory of Elastic Wave Scattering: Applications of the Method of Optimal Truncation," *Journal of Applied Physics*, 58 (3), 1985, pp. 1102-1115.
- A-9. A. Minachi, F.J. Margetan, and D.K. Hsu, "Delamination Sizing in Composite Materials Using a Gauss-Hermite Beam Model," *Ultrasonics*, 31 (4), 1993, pp. 237-243.
- A-10. C.-P. Chiou, F.J. Margetan, and R.B. Thompson, "Modeling of Ultrasonic Signals From Weak Inclusions," in *Review of Progress in Quantitative NDE*, Vol. 15A, D.O. Thompson and D. E. Chimenti, eds., Plenum Press, NY, pp. 49-55, 1996.
- A-11. C.-P. Chiou, F.J. Margetan, and R.B. Thompson, "Development of Ultrasonic Models for Hard-Alpha Inclusions in Titanium Alloys," in *Review of Progress in Quantitative NDE*, Vol. 16B, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1529-1536, 1997.

- A-12. J.E. Gubernatis, E. Domany, J.A. Krumhansl, and M. Huberman, "The Born Approximation in the Theory of the Scattering of Elastic Waves by Flaws," *Journal of Applied Physics*, 48 (7), 1977, pp. 2812-2819.
- A-13. J.H. Rose and J.M. Richardson, "Time Domain Born Approximation," *Journal of Nondestr. Eval.* 3 (1), 1982, pp. 45-53.
- A-14. L.W. Schmerr, Jr., *Fundamentals of Ultrasonic Nondestructive Evaluation – A Modeling Approach*, Kluwer Academic Publishers, June 1998.
- A-15. M.F.X. Gigliotti, R.S. Gilmore, and L.C. Perocchi, "Microstructure and Sound Velocity of Ti-N-O Synthetic Inclusions in Ti-6Al-4V," *Metallurgical and Materials Transactions A*, Vol. 25A, pp. 2321-2329, 1994.
- A-16. C.F. Ying and R. Truell, "Scattering of a Plane Longitudinal Wave by a Spherical Obstacle in an Isotropically Elastic Solid," *Journal of Applied Physics*, 27, 1956, pp. 1086-1097.
- A-17. C.-P. Chiou, B. Boyd, R.B. Thompson, and J. Oliver, "Geometrical and Ultrasonic Modeling of Hard-Alpha Defect Components in Titanium," in *Review of Progress in Quantitative NDE*, Vol. 18, D.O. Thompson and D.E. Chimenti, eds., Plenum Press, NY, pp. 1829-1836, 1998.